Digital Communication Systems

S Le Goff

October 2023

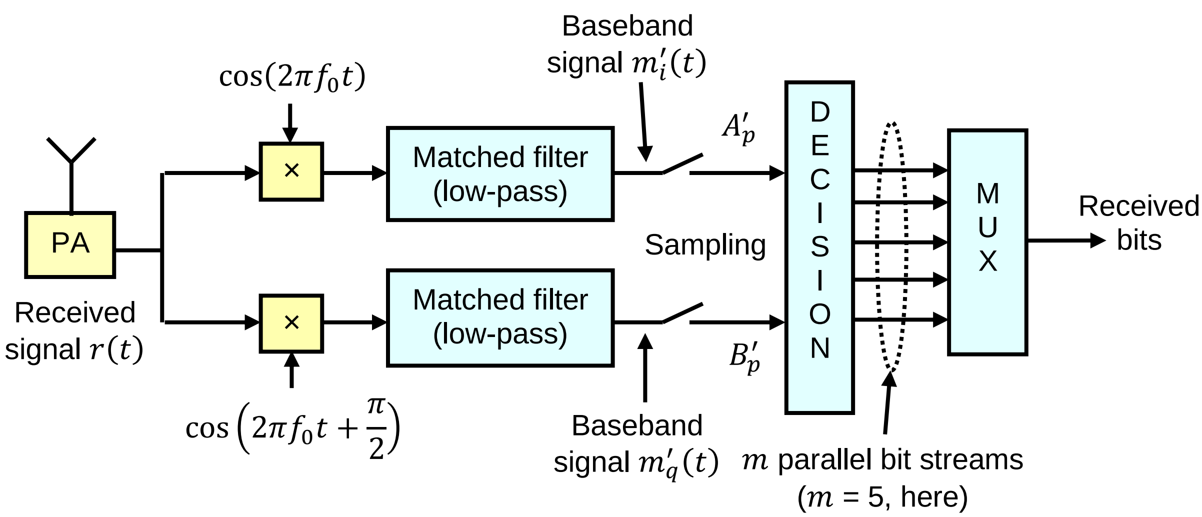
School of Engineering

Newcastle University

Contents

8. Digital Communication Receivers

We have represented below the generic structure of a digital communication receiver.



The received signal must be amplified since the signal power received by the antenna is usually quite low.

Some thermal noise is added to the received signal at this stage. This noise originates from the movement of electrons inside the connections and components of the RF power amplifier used at the front-end of the radio receiver, often referred to as a *low-noise amplifier*.

It is necessary for a low-noise amplifier to boost the desired signal power, i.e., amplify $s(t)$, while adding as little noise and distortion as possible so that the retrieval of the transmitted symbols C_k is possible in the later stages of the receiver. This is in fact at that stage that a noise process $n(t)$ is added to the signal $s(t)$. This noise added by the low-noise amplifier is modeled as *additive white Gaussian noise* (AWGN). The temporal and spectral characteristics of a white and Gaussian noise process were explained in previous chapters.

In the real world, the AWGN is not the only disturbance that affects the transmitted signal $s(t)$. For instance, many wireless channels are subject to interference that is even more detrimental to communications integrity than the AWGN:

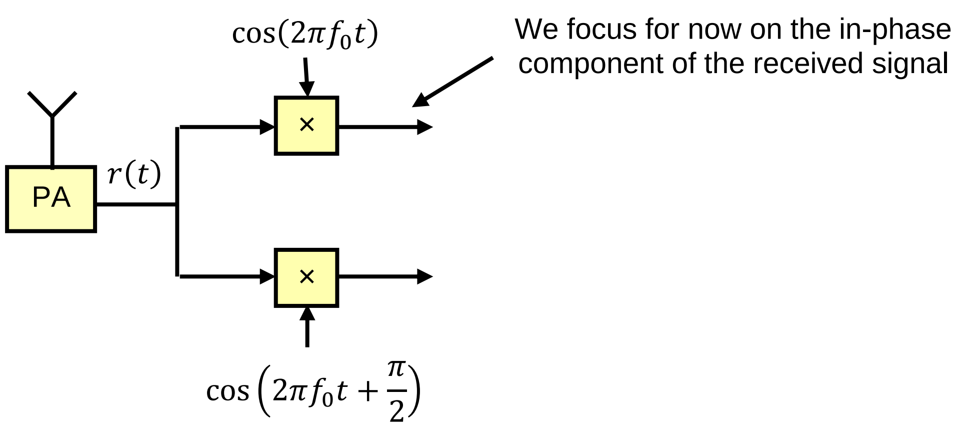
- Interference among different users. We talk here about phenomena such as multi-user interference and co-channel interference that are beyond the scope of this module.
- Interference among several randomly-attenuated and randomly-delayed versions of the same signal. This kind of interference is typically present on multi-path channels.
- Interference among successive symbols carried by the same signal. This is known as *inter-symbol interference*. We already had a taste of this issue in the previous chapter.

Throughout this chapter, we will assume for now that the channel is an AWGN channel, i.e., the received signal $r(t)$ at the low-noise amplifier output is given by

$$r(t) = s(t) + n(t),$$

where $s(t)$ is the transmitted band-pass signal and $n(t)$ is a zero-mean Gaussian noise process with constant power spectral density (PSD), i.e., $\Phi_n(f) = \frac{N_0}{2}$ for all frequency values. Note that we do not define a variance for $n(t)$ as the variance of any white random process is infinite.

- **Signal present at the in-phase mixer output**



First, we are going to focus on the in-phase component of the receiver.

To obtain the base-band signal $m_i'(t)$, which is an estimate of the base-band signal $m_i(t)$ present at the transmitter side, one has first to demodulate the signal $r(t)$ by mixing it with a locally-generated carrier signal $\cos(2\pi f_0 t)$.

The expression of the received signal can be written as follows:

$$r(t) = s(t) + n(t) = \sum_{k=0}^{+\infty} A_k \cdot h(t-kT) \cdot \cos(2\pi f_0 t) + B_k \cdot h(t-kT) \cdot \cos\left(2\pi f_0 t + \frac{\pi}{2}\right) + n(t),$$

which is equivalent to

$$r(t) = s(t) + n(t) = \sum_{k=0}^{+\infty} A_k \cdot h(t-kT) \cdot \cos(2\pi f_0 t) - B_k \cdot h(t-kT) \cdot \sin(2\pi f_0 t) + n(t).$$

The signal at the mixer output is then expressed as

$$r(t) \cdot \cos(2\pi f_0 t) = \sum_{k=0}^{+\infty} A_k \cdot h(t-kT) \cdot \cos^2(2\pi f_0 t) - B_k \cdot h(t-kT) \cdot \sin(2\pi f_0 t) \cdot \cos(2\pi f_0 t) + n(t) \cdot \cos(2\pi f_0 t).$$

We remember that

$$\cos(a)\cos(b) = \frac{1}{2} \cdot [\cos(a+b) + \cos(a-b)] \text{ and } \sin(a)\cos(b) = \frac{1}{2} \cdot [\sin(a+b) + \sin(a-b)].$$

We thus have $\cos^2(2\pi f_0 t) = \frac{1}{2} \cdot [1 + \cos(4\pi f_0 t)]$ and $\cos(2\pi f_0 t) \cdot \sin(2\pi f_0 t) = \frac{1}{2} \cdot \sin(4\pi f_0 t)$.

Therefore, the signal at the mixer output can be written as

$$r(t) \cdot \cos(2\pi f_0 t) = \sum_{k=0}^{+\infty} \frac{A_k}{2} \cdot h(t-kT) + \sum_{k=0}^{+\infty} \frac{A_k}{2} \cdot h(t-kT) \cdot \cos(4\pi f_0 t) - \sum_{k=-\infty}^{+\infty} \frac{B_k}{2} \cdot h(t-kT) \cdot \sin(4\pi f_0 t) + n(t) \cdot \cos(2\pi f_0 t).$$

We now need to find the characteristics of the noise $n_i'(t) = n(t) \cdot \cos(2\pi f_0 t)$ present at the in-phase mixer output. We must remember that, at any time t, $n(t)$ is a random variable as $n(t)$ is a random process, whereas $\cos(2\pi f_0 t)$ is simply a number as $\cos(2\pi f_0 t)$ is a deterministic signal.

The noise process $n_i'(t)$ has the same distribution as $n(t)$ since multiplying a random variable by any number does not have any effect on the shape of its probability density function.

Recall that $n(t)$ is Gaussian with a mean equal to zero. Therefore, $n_i'(t)$ is also Gaussian with a mean m given by

$$m = E\{n_i'(t)\} = E\{n(t)\} \cdot E\{\cos(2\pi f_0 t)\} = E\{n(t)\} \cdot \cos(2\pi f_0 t) = 0.$$

We also need to know whether $n_i'(t)$ is a white noise process. One way to answer this question consists of determining its autocorrelation function. The autocorrelation function $\Gamma_{n_i'}(t)$ of the random process $n_i'(t)$ can be computed as follows:

$$\Gamma_{n_i'}(t) = E_\tau\{n_i'(\tau) \cdot n_i'(\tau - t)\} = E_\tau\{n(\tau) \cdot n(\tau - t)\} \cdot E_\tau\{\cos(2\pi f_0 \tau) \cdot \cos(2\pi f_0 [\tau - t])\}$$

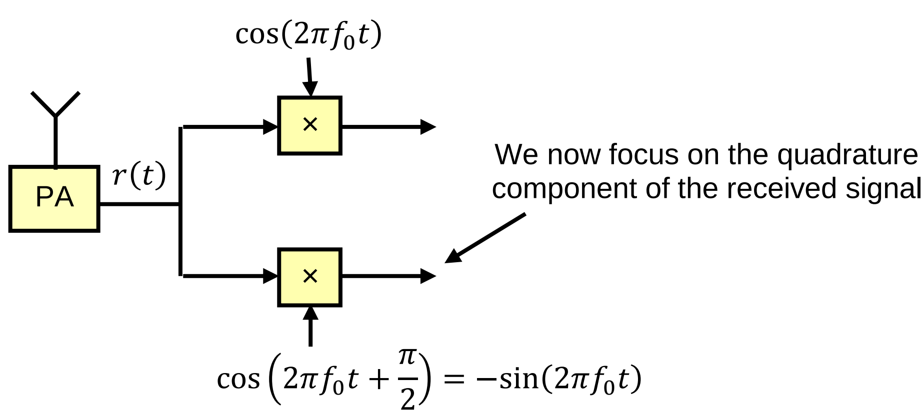
$$= \frac{1}{2} \cdot \Gamma_n(t) \cdot [E_\tau\{\cos(2\pi f_0 t)\} + E_\tau\{\cos(2\pi f_0 [2\tau - t])\}] = \frac{1}{2} \cdot \Gamma_n(t) \cdot \cos(2\pi f_0 t)$$

$$= \frac{N_0}{4} \cdot \delta(t) \cdot \cos(2\pi f_0 t) = \frac{N_0}{4} \cdot \delta(t) \cdot \cos(0) = \frac{N_0}{4} \cdot \delta(t).$$

This is the autocorrelation function of a white noise process with a power spectral density $\Phi_{n_i'}(f) = \frac{N_0}{4}$ for all frequencies. Hence, $n_i'(t)$ is a white random process.

• **Signal present at the quadrature mixer output**

We are now going to focus on the quadrature component of the receiver.



To obtain the base-band signal $m'_q(t)$, which is an estimate of the base-band signal $m_q(t)$, present at the transmitter side, one has first to demodulate the signal $r(t)$ by mixing it with a locally-generated carrier signal $\cos\left(2\pi f_0 t + \frac{\pi}{2}\right) = -\sin(2\pi f_0 t)$.

Recall that the expression of the received signal is given by

$$r(t) = s(t) + n(t) = \sum_{k=0}^{+\infty} A_k \cdot h(t - kT) \cdot \cos(2\pi f_0 t) - B_k \cdot h(t - kT) \cdot \sin(2\pi f_0 t) + n(t).$$

The signal at the quadrature mixer output is then expressed as

$$-r(t) \cdot sin(2\pi f_0 t) = \sum_{k=0}^{+\infty} -A_k \cdot h(t - kT) \cdot \cos(2\pi f_0 t) \cdot \sin(2\pi f_0 t) + B_k \cdot h(t - kT) \cdot \sin^2(2\pi f_0 t) - n(t) \cdot \sin(2\pi f_0 t).$$

We remember that

$\cos(a)\sin(b) = \frac{1}{2} \cdot [\sin(a + b) - \sin(a - b)]$ and $\sin(a)\sin(b) = \frac{1}{2} \cdot [\cos(a - b) - \cos(a + b)]$.

We thus have $\cos(2\pi f_0 t)\sin(2\pi f_0 t) = \frac{1}{2} \cdot \sin(4\pi f_0 t)$ and $\sin^2(2\pi f_0 t) = \frac{1}{2} \cdot [1 - \cos(4\pi f_0 t)]$.

We thus obtain

$$-r(t) \cdot \sin(2\pi f_0 t) = -\sum_{k=0}^{+\infty} \frac{A_k}{2} \cdot h(t - kT) \cdot \sin(4\pi f_0 t) + \sum_{k=0}^{+\infty} \frac{B_k}{2} \cdot h(t - kT) - \sum_{k=0}^{+\infty} \frac{B_k}{2} \cdot h(t - kT) \cdot \cos(4\pi f_0 t) - n(t) \cdot \sin(2\pi f_0 t).$$

We now need to find the characteristics of the noise $n'_q(t) = -n(t) \cdot \sin(2\pi f_0 t)$ present at the quadrature mixer output. We must remember that, at any time t, $n(t)$ is a random variable as $n(t)$ is a random process, whereas $\sin(2\pi f_0 t)$ is simply a number as $\sin(2\pi f_0 t)$ is a deterministic signal.

The noise process $n'_q(t)$ has the same distribution as the noise process $n(t)$ since multiplying a random variable by any number does not have any effect on the shape of its probability density function.

Recall that $n(t)$ is Gaussian with a mean equal to zero. Therefore, $n'_q(t)$ is also Gaussian with a mean m given by

$$m = E\{n'_q(t)\} = -E\{n(t)\} \cdot E\{\sin(2\pi f_0 t)\} = -E\{n(t)\} \cdot \sin(2\pi f_0 t) = 0.$$

We also need to know whether $n'_q(t)$ is a white noise process. To do so, we are going to determine its autocorrelation function.

The autocorrelation function $\Gamma_{n'_q}(t)$ of the random process $n'_q(t) = -n(t) \cdot \sin(2\pi f_0 t)$ can be computed as follows:

$$\Gamma_{n'_q}(t) = E_\tau\{n'_q(\tau) \cdot n'_q(\tau - t)\} = E_\tau\{n(\tau) \cdot n(\tau - t)\} \cdot E_\tau\{\sin(2\pi f_0 \tau) \cdot \sin(2\pi f_0 [\tau - t])\}$$

$$= \frac{1}{2} \cdot \Gamma_n(t) \cdot [E_\tau\{\cos(2\pi f_0 t)\} - E_\tau\{\cos(2\pi f_0 [2\tau - t])\}] = \frac{1}{2} \cdot \Gamma_n(t) \cdot \cos(2\pi f_0 t)$$

$$= \frac{N_0}{4} \cdot \delta(t) \cdot \cos(2\pi f_0 t) = \frac{N_0}{4} \cdot \delta(t) \cdot \cos(0) = \frac{N_0}{4} \cdot \delta(t).$$

This is the autocorrelation function of a white noise process with a power spectral density $\Phi_{n'_q}(f) = \frac{N_0}{4}$ for all frequencies. Hence, $n'_q(t)$ is a white random process.

• **Are $n'_i(t)$ and $n'_q(t)$ independent random processes?**

To answer this question, we must first realise that the random processes $n'_i(t)$ and $n'_q(t)$ are jointly Gaussian. In other words, any linear combination of them is Gaussian.

It is indeed easy to see that a random process defined as

$$\alpha \cdot n'_i(t) + \beta \cdot n'_q(t) = n(t) \cdot [\alpha \cdot \cos(2\pi f_0 t) - \beta \cdot \sin(2\pi f_0 t)],$$

where α and β designate two arbitrary real numbers, is Gaussian.

If two jointly Gaussian random processes are uncorrelated, then they are independent. So, to prove the independence of $n'_i(t)$ and $n'_q(t)$, we now need to show that these two jointly Gaussian random processes are uncorrelated.

The correlation between the random processes $n_i'(t)$ and $n_q'(t)$ can be determined by evaluating their cross-correlation function, defined as $\Gamma_{n_i',n_q'}(t) = E_\tau\{n_i'(\tau)\cdot n_q'(\tau - t)\}$, and showing that it is equal to $E_\tau\{n_i'(\tau)\}\cdot E_\tau\{n_q'(\tau - t)\} = 0$. As $n_i'(t)$ and $n_q'(t)$ have means equal to zero, we indeed have $E_\tau\{n_i'(\tau)\} = E_\tau\{n_q'(\tau - t)\} = 0$.

The cross-correlation function $\Gamma_{n_i',n_q'}(t)$ can be computed as follows:

$$\Gamma_{n_i',n_q'}(t) = E_\tau\{n_i'(\tau)\cdot n_q'(\tau - t)\} = -E_\tau\{n(\tau)\cdot n(\tau - t)\}\cdot E_\tau\{\cos(2\pi f_0\tau)\cdot sin(2\pi f_0[\tau - t])\}$$

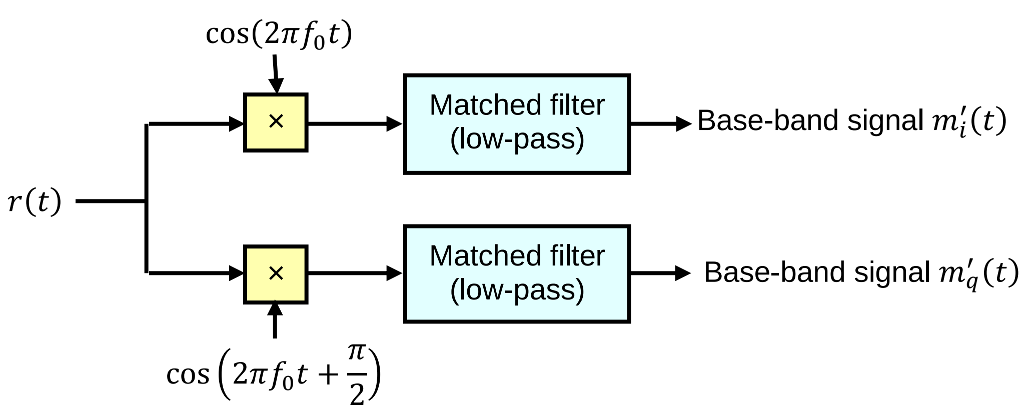
$$= -\frac{1}{2}\cdot\Gamma_n(t)\cdot[E_\tau\{\sin(2\pi f_0[2\tau - t])\} - E_\tau\{\sin(2\pi f_0 t)\}] = \frac{1}{2}\cdot\Gamma_n(t)\cdot\sin(2\pi f_0 t)$$

$$= \frac{N_0}{4}\cdot\delta(t)\cdot\sin(2\pi f_0 t) = \frac{N_0}{4}\cdot\delta(t)\cdot\sin(0) = 0.$$

As the cross-covariance function $\Gamma_{n_i',n_q'}(t)$ is equal to zero for all values of t, we can state that the random processes $n_i'(t)$ and $n_q'(t)$ are uncorrelated. As they are also jointly Gaussian, we then conclude that they are independent.

• **Expression of the in-phase baseband signal $m_i'(t)$**

Both signals present at the mixer outputs then go through low-pass filters with an impulse response $q(t)$ and a transfer function $Q(f)$.



Let us first focus on the in-phase component. At the in-phase low-pass filter output, we obtain the base-band signal $m_i'(t)$ given by

$$m_i'(t) = \sum_{k=0}^{+\infty} \frac{A_k}{2} \cdot h(t-kT) * q(t) + n_i'(t) * q(t).$$

But, this low-pass filter is in fact designed very carefully. In order to maximize the signal-to-noise ratio (SNR) of the base-band signal, the low-pass filter is designed so that its impulse response $q(t)$ is *matched* to the pulse shape. The demonstration of this result is not detailed here as it is a rather lengthy one that is more suited to a module on detection theory.

In other words, the impulse response $q(t)$ is given by $q(t) = h\left(T_p - t\right)$. The use of a *matched filter* at the receiver side guarantees optimal error performance at the receiver output.

We can thus write $m_i'(t) = \sum_{k=0}^{+\infty} \frac{A_k}{2} \cdot h(t-kT) * h\left(T_p - t\right) + n_i'(t) * h\left(T_p - t\right)$.

Let us focus first on the first term in this expression.

It can be shown that the Fourier transform of the term $h(t-kT) * h\left(T_p - t\right)$ is expressed as

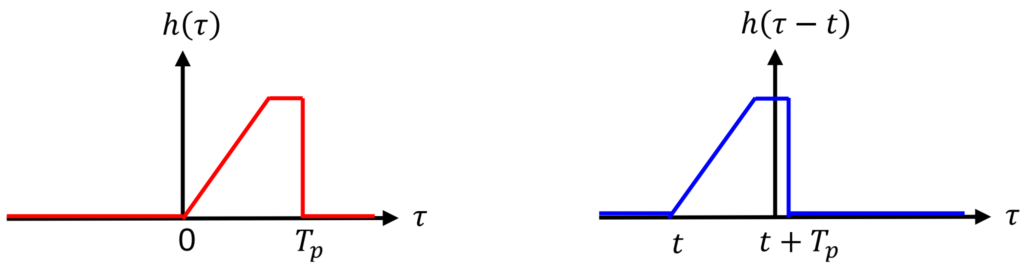
$$H(f) \cdot e^{-j2\pi fkT} \cdot H^*(f) \cdot e^{-j2\pi fT_p} = H(f) \cdot H^*(f) \cdot e^{-j2\pi f\left(kT+T_p\right)} = G(f) \cdot e^{-j2\pi f\left(kT+T_p\right)}.$$

The result is the Fourier transform of a term $g\left(t-kT-T_p\right)$, with $g(t) = h(t) * h(-t) = \int_{-\infty}^{+\infty} h(\tau) \cdot h(\tau - t) d\tau$.

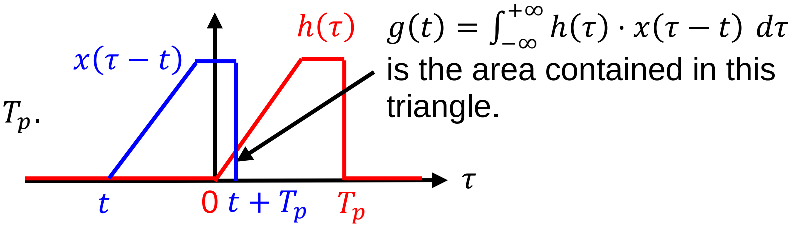
Note that $g(0) = \int_{-\infty}^{+\infty} [h(\tau)]^2 \, d\tau = E_h$, where E_h designates the energy of the pulse $h(t)$.

We then obtain $m_i'(t) = \sum_{k=0}^{+\infty} \frac{A_k}{2} \cdot g\left(t-kT-T_p\right) + n_i'(t) * h\left(T_p - t\right)$.

The fact that the pulse $h(t)$ has been replaced by another pulse $g(t)$ in the signal component merely reflects the distortion effect due to the low-pass filter.



We have
$g(t) = 0$ for $t < -T_p$ or $t > T_p$.
$g(t)$ is maximal for $t = 0$.



Let us now determine the characteristics of the noise process $n_i''(t) = n_i'(t) * h(T_p - t)$ present at the matched filter output.

The noise process $n_i''(t)$ is not white because of the presence of a low-pass filter. However, $n_i''(t)$ is Gaussian since linear filtering of a Gaussian process produces another Gaussian process.

The mean of $n_i''(t)$ is computed as follows:

$$E\{n_i''(t)\} = E\{n_i'(t) * h(T_p - t)\} = E\left\{\int_{-\infty}^{+\infty} n_i'(\tau) \cdot h(T_p + \tau - t) d\tau\right\}$$

$= \int_{-\infty}^{+\infty} E\{n_i'(\tau) \cdot h(T_p + \tau - t)\} d\tau = \int_{-\infty}^{+\infty} E\{n_i'(\tau)\} \cdot h(T_p + \tau - t) d\tau = 0$, as $E\{n_i'(\tau)\} = 0$.

The variance of $n_i''(t)$ is not infinite because $n_i''(t)$ is not a white noise process. We thus need to evaluate the variance of $n_i''(t)$. To do so, one can start with the definition of the variance of a random process: $\sigma^2 = E\{[n_i''(t)]^2\} - [E\{n_i''(t)\}]^2 = E\{[n_i''(t)]^2\}$ as $E\{n_i''(t)\} = 0$.

As the term $E\{[n_i''(t)]^2\}$ is in fact the value of the autocorrelation function $\Gamma_{n_i''}(t)$ of $n_i''(t)$ at time $t = 0$, we have $\sigma^2 = \Gamma_{n_i''}(0)$.

We also know that $\Gamma_{n_i''}(t)$ is the inverse Fourier transform of the power spectral density $\Phi_{n_i''}(f)$. Therefore, we can write $\sigma^2 = \Gamma_{n_i''}(0) = \int_{-\infty}^{+\infty} \Phi_{n_i''}(f)df$.

We can now use the expression giving the PSD of a random process at the output of a linear filter with a transfer function $H^*(f) \cdot e^{-j2\pi f T_p}$:

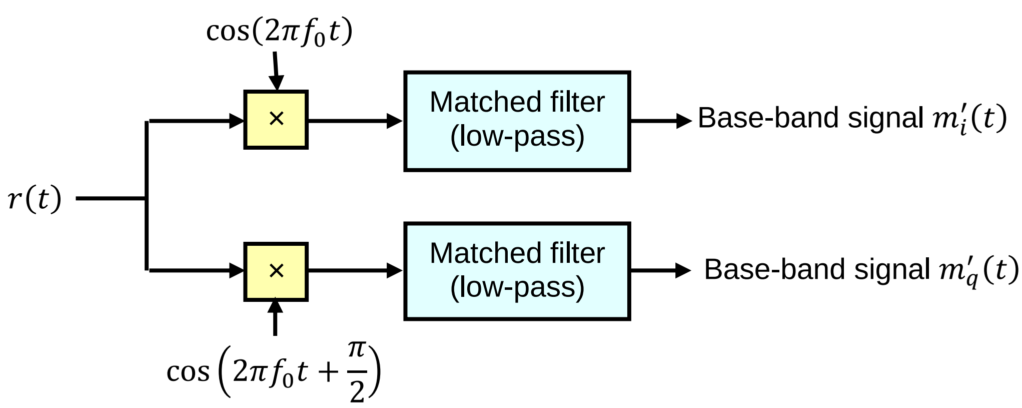
$$\Phi_{n_i''}(f) = \Phi_{n_i'}(f) \cdot \left|H^*(f) \cdot e^{-j2\pi f T_p}\right|^2 = \Phi_{n_i'}(f) \cdot |H(f)|^2 = \frac{N_0}{4} \cdot G(f).$$

We thus have $\sigma^2 = \int_{-\infty}^{+\infty} \Phi_{n_i''}(f)df = \frac{N_0}{4} \cdot \int_{-\infty}^{+\infty} G(f)df$.

As $G(f)$ is the Fourier transform of $g(t)$, it is easy to show that the term $\int_{-\infty}^{+\infty} G(f)df$ is simply the value of $g(t)$ at time $t = 0$: $g(0) = \int_{-\infty}^{+\infty} G(f)df$.

This finally leads to the expression $\sigma^2 = \frac{N_0}{4} \cdot g(0) = \frac{N_0 E_h}{4}$, where E_h designates the energy of the pulse $h(t)$.

• **Expression of the quadrature baseband signal $m_q'(t)$**



Like its in-phase counterpart, the signal present at the quadrature mixer output goes through a low-pass matched filter with an impulse response $h(T_p - t)$ and a transfer function $H^*(f) \cdot e^{-j2\pi f T_p}$.

By using the same derivation as above, we obtain the expression of the base-band signal $m'_q(t)$:

$$m'_q(t) = \sum_{k=0}^{+\infty} \frac{B_k}{2} \cdot g(t - kT - T_p) + n''_q(t),$$

where $n''_q(t)$ is a zero-mean Gaussian noise process with a variance $\sigma^2 = \frac{N_0 E_h}{4}$.

- **Expression of the complex baseband signal $\boldsymbol{m'(t) = m'_i(t) + jm'_q(t)}$**

It is possible to combine $m'_i(t)$ and $m'_q(t)$ into a single complex baseband signal $m'(t)$ whose expression is

$$m'(t) = m'_i(t) + jm'_q(t) = \sum_{k=0}^{+\infty} \frac{1}{2} C_k g(t - kT - T_p) + n''(t).$$

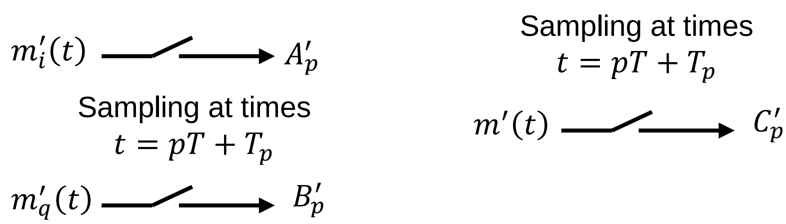
where $C_k = A_k + jB_k$ is the complex symbol transmitted at time kT and $n''(t) = n''_i(t) + jn''_q(t)$ designates a complex zero-mean Gaussian noise process with a variance $2\sigma^2 = \frac{N_0 E_h}{2}$, meaning that the variances of $n''_i(t)$ and $n''_q(t)$ are both given by $\sigma^2 = \frac{N_0 E_h}{4}$.

The noise processes $n''_i(t)$ and $n''_q(t)$ are independent.

From now on, we are going to adopt this notation as it is not only very elegant but also allows us to jointly process both in-phase and quadrature components.

- **Sampling of the complex baseband signal $\boldsymbol{m'(t)}$**

The complex baseband signal $m'(t)$ is then sampled at times $t = pT + T_p$, where p designates an integer. By doing so, we can obtain an estimate $C'_p = A'_p + jB'_p$ of the complex symbol $C_p = A_p + jB_p$ that was transmitted at time $t = pT$.



By sampling $m'(t)$ at time $t = pT + T_p$, we obtain a sample C_p' which is expressed as follows:

$$C_p' = m'\left(pT + T_p\right) = \sum_{k=0}^{+\infty} \frac{1}{2} C_k g\left((p-k)T\right) + n_p,$$

where $n_p = n''\left(pT + T_p\right)$ designates a complex zero-mean Gaussian noise sample with a variance $2\sigma^2 = \frac{N_0 E_h}{2}$.

To better understand this equation, let us develop it:

$$C_p' = \cdots + \frac{C_{p-2}}{2} g(2T) + \frac{C_{p-1}}{2} g(T) + \frac{C_p}{2} g(0) + \frac{C_{p+1}}{2} g(-T) + \frac{C_{p+2}}{2} g(-2T) + \cdots + n_p,$$

This equation indicates that a sample C_p' depends on several consecutive transmitted symbols ... $C_{p-2}, C_{p-1}, C_p, C_{p+1}, C_{p+2}$ .... In fact, this dependency extends over a limited number of symbols because we have $g(t) = 0$ for $t < -T_p$ and $t > T_p$.

For instance, if the pulse duration T_p is greater than the symbol duration T, but less than $2T$, i.e., is such that $T < T_p < 2T_p$, we can then write $C_p' = \frac{C_{p-1}}{2} g(T) + \frac{C_p}{2} g(0) + \frac{C_{p+1}}{2} g(-T) + n_p$.

In any case, we can see that the sample C_p' does not only depend on the symbol C_p, but also on other symbols C_k, $k \neq p$. The presence of such *inter-symbol interference* (ISI) may lead to very severe, and generally unacceptable, degradation of the error performance at the receiver output.

In practice, ISI must thus be avoided at any cost. If ISI cannot be avoided, such as in the case of frequency-selective fading channels that will be studied later, then it must be cancelled at the receiver side using techniques such as equalization or orthogonal frequency division multiplexing (OFDM).

Here, in the case of an AWGN channel, if the pulse $g(t)$ is carefully chosen so that $g(0) \neq 0$ and $g\big((p-k)T\big) = 0$ for any $k \neq p$, then ISI is completely avoided, and we then obtain

$$C'_p = \frac{C_p}{2} g(0) + n_p = \frac{E_h}{2} C_p + n_p,$$

where $E_h = g(0)$ designates the energy of the pulse $h(t)$.

• **Raised-cosine pulses**

The condition $g(0) \neq 0$ and $g\big((p-k)T\big) = 0$ for any $k \neq p$ is termed *Nyquist criterion*. The range of pulses $g(t)$, for which $g(0) \neq 0$ and $g\big((p-k)T\big) = 0$ for any $k \neq p$, constitutes a family of pulses called *raised-cosine pulses*. In other words, raised-cosine pulses satisfy the Nyquist criterion, and their use guarantees ISI-free transmission in the case of the AWGN channel.

The equation of a raised-cosine pulse is in the form

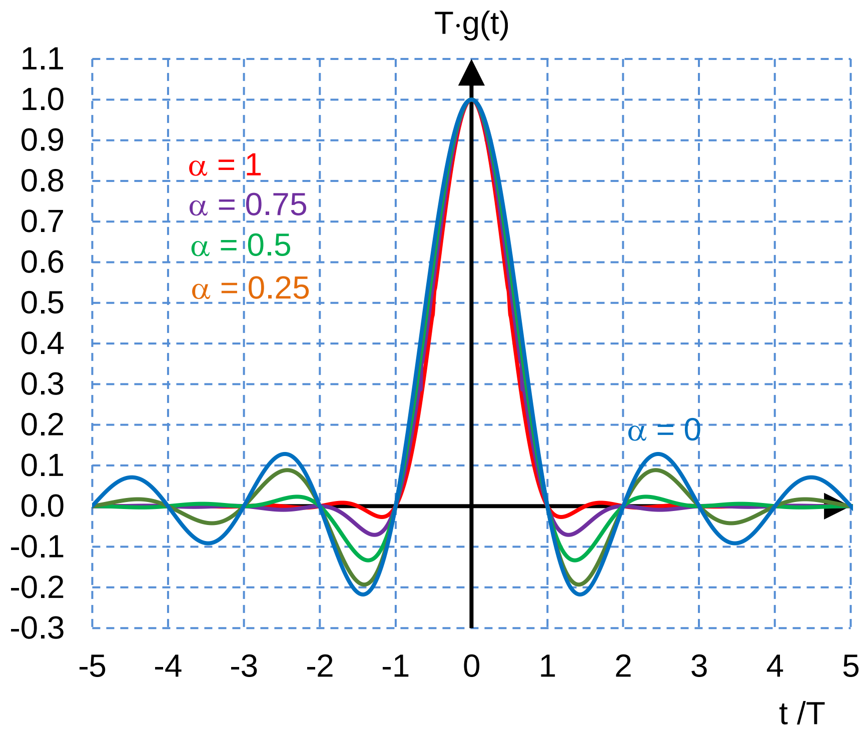
$$g(t) = \frac{1}{T} \cdot \operatorname{sinc}\left(\frac{\pi t}{T}\right) \cdot \frac{\cos\left(\frac{\pi \alpha t}{T}\right)}{1 - \frac{4\alpha^2 t^2}{T^2}},$$

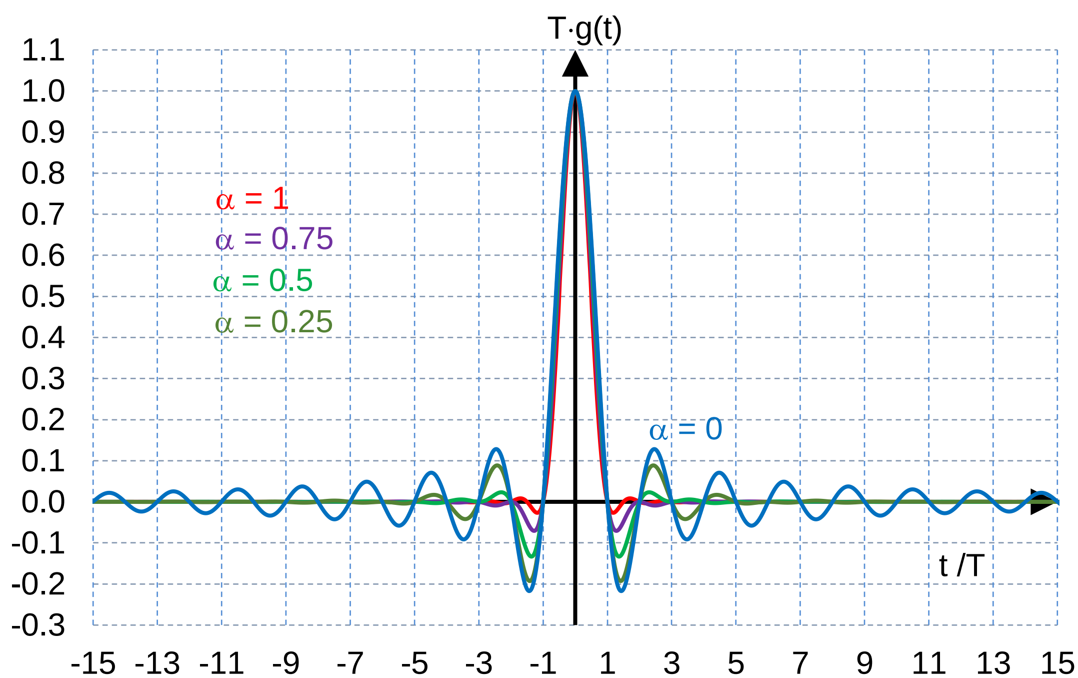
where the parameter α, called *roll-off factor*, can take any value in the range from 0 to 1. This is the system parameter that was introduced earlier.

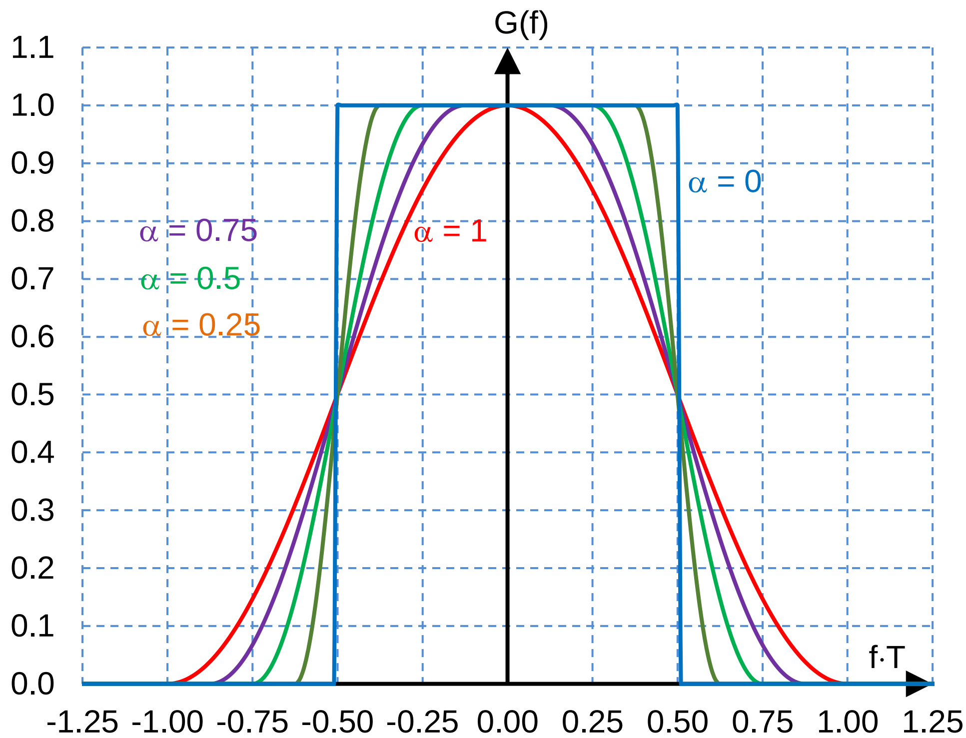
The Fourier transform of a raised-cosine pulse is in the form

$$G(f) = \begin{cases} 1 & \text{for} \quad f \leq \frac{1-\alpha}{2T} \\ \frac{1}{2} \cdot \left[1 + \cos\left(\frac{\pi T}{\alpha}\left(f - \frac{1-\alpha}{2T}\right)\right)\right] & \text{for} \quad \frac{1-\alpha}{2T} < f \leq \frac{1+\alpha}{2T}. \\ 0 & \text{for} \quad f > \frac{1+\alpha}{2T} \end{cases}$$

This equation shows that the bandwidth of a raised cosine pulse is equal to $\frac{1+\alpha}{2T}$, and thus ranges from $\frac{1}{2T}$, for $\alpha = 0$, to $\frac{1}{T}$, for $\alpha = 1$.


T·g(t)
1.1
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
α = 1
α = 0.75
α = 0.5
α = 0.25
α = 0
-5
-4
-3
-2
-1
0
1
2
3
4
5
t /T


T·g(t)
1.1
1.0
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
-0.1
-0.2
-0.3
α = 1
α = 0.75
α = 0.5
α = 0.25
α = 0
t /T
-15
-13
-11
-9
-7
-5
-3
-1
1
3
5
7
9
11
13
15



• Bandwidth of the base-band and modulated signals

In the previous chapter, we saw that the power spectral density (PSD), $M(f)$, of the base-band signal $m(t)$ is given by the following expression:

$$M(f) = K \cdot |H(f)|^2,$$

where K is a constant and $H(f)$ is the Fourier transform of the pulse $h(t)$ that carries the symbols C_k.

We have also seen in this chapter that we have $G(f) = |H(f)|^2$. Therefore, we can now write

$$M(f) = K \cdot G(f).$$

This expression clearly shows that the bandwidth B_{bb} of $M(f)$ is equal to that of $G(f)$:

$$B_{bb} = \frac{1+\alpha}{2T}.$$

Hence, the bandwidth B_{bp} of the modulated signal $s(t)$ is given by

$$B_{bp} = 2B_{bb} = \frac{1+\alpha}{T}.$$

The case "$\alpha = 0$" is optimal in terms of bandwidth but the corresponding pulse shape is difficult to implement. The case "$\alpha = 1$" is the simplest to implement, but it requires the largest bandwidth.

In any case, we conclude that ISI-free transmission over an AWGN channel requires a bandwidth of, at least, $\frac{1}{2T}$ for the base-band signal and, at least, $\frac{1}{T}$ for the band-pass (modulated) signal.

• Root raised cosine pulses

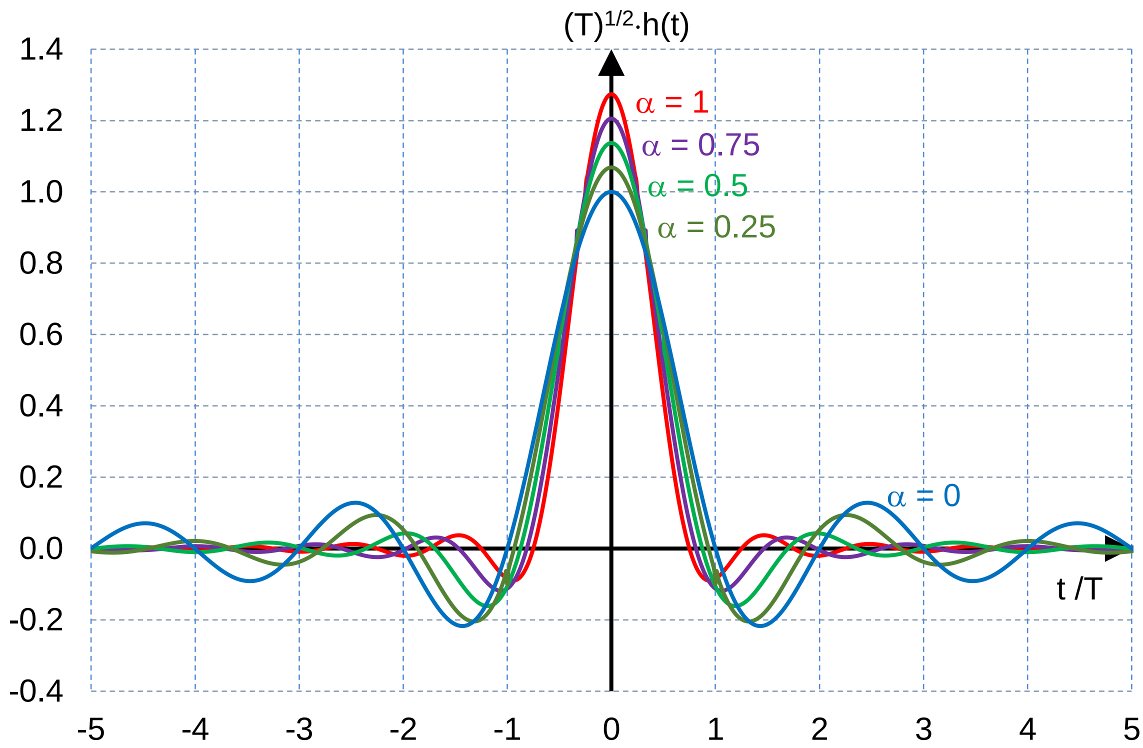
We can now determine the pulse shape $h(t)$ that should be used to carry the symbols C_k. Since we have $G(f) = |H(f)|^2$, we can simply choose the pulse shape so that its Fourier transform is given by $H(f) = \sqrt{G(f)}$, assuming that $H(f)$ is a real function.

Therefore, the pulse $h(t)$ should be designed so that $H(f)$ is the square-root of the Fourier transform of a raised-cosine pulse. A lengthy derivation would lead us to the following expression of $h(t)$:

$$h(t) = \frac{1}{\sqrt{T}} \cdot \frac{\sin\left(\pi(1-\alpha)\frac{t}{T}\right) + 4\alpha\frac{t}{T} \cdot \cos\left(\pi(1+\alpha)\frac{t}{T}\right)}{\pi\frac{t}{T} \cdot \left(1 - 4\alpha^2\left(\frac{t}{T}\right)^2\right)}.$$

The corresponding pulse is shown below for five different values of the roll-off factor α. We now know what pulse shapes should be used at the transmitter side.

Note that the five pulses are displayed as centered around $t = 0$. In practice, they should of course be shifted so that $h(t) = 0, \forall t < 0$, in order to make them causal.



• AWGN channel model and signal-to-noise ratio

We have seen that the receiver is able to produce an estimate C'_p of the complex symbol C_p that was transmitted at time $t = pT$. Over an AWGN channel, this estimate is given by the equation

$$C'_p = \frac{E_h}{2} C_p + n_p,$$

where E_h designates the energy of the pulse $h(t)$ used to carry symbols C_p over the physical channel, and n_p denotes a zero-mean complex Gaussian sample with variance $2\sigma^2 = \frac{N_0 E_h}{2}$.

We can further simplify the expression of C'_p by multiplying it by a factor $\frac{2}{E_h}$, which does not change anything in practice of course. We now obtain a very simple expression for the estimate C'_p of the complex symbol C_p that was transmitted at time $t = pT$:

$$C'_p = C_p + n_p,$$

where n_p now denotes a zero-mean complex Gaussian sample with variance $2\sigma^2 = \left(\frac{2}{E_h}\right)^2 \frac{N_0 E_h}{2} = \frac{2N_0}{E_h}$.

At this stage, we would like to introduce the concept of signal-to-noise ratio (SNR). This metric provides a measure of the strength of the received signal with respect to that of the white Gaussian noise process.

We have previously seen that the expression of the received signal, i.e., the signal $r(t)$ available at the output of the RF low-noise power amplifier, is

$$r(t) = s(t) + n(t) = \sum_{k=0}^{+\infty} A_k \cdot h(t-kT) \cdot \cos(2\pi f_0 t) - B_k \cdot h(t-kT) \cdot \sin(2\pi f_0 t) + n(t).$$

The strength of the signal component in $r(t)$ is measured by the average energy, E_s, allocated to each symbol $C_k = A_k + jB_k$, whereas the strength of the noise component in $r(t)$ is measured by the amplitude of the power spectral density $\Phi_n(f)$ of the noise process $n(t)$. As we have $\Phi_n(f) = \frac{N_0}{2}, \forall f$, the strength of the noise component can thus simply be measured by the parameter N_0.

For now, the signal-to-noise ratio is thus going to be defined as the ratio $\frac{E_s}{N_0}$. The latter is referred to as the *average SNR per symbol*.

The average energy, E_s, allocated to each complex symbol C_k can be computed as follows:

$$E_s = E_{A_k,B_k}\left\{\int_{-\infty}^{+\infty}\left(A_k h(t-kT)\cos(2\pi f_0 t) - B_k h(t-kT)\sin(2\pi f_0 t)\right)^2 dt\right\}.$$

By developing this expression, we obtain

$$E_s = E_{A_k,B_k}\left\{\int_{-\infty}^{+\infty} A_k^2 h(t-kT)^2\cos^2(2\pi f_0 t) + B_k^2 h(t-kT)^2\sin^2(2\pi f_0 t) - 2A_k B_k h(t-kT)^2\cos(2\pi f_0 t)\sin(2\pi f_0 t)\,dt\right\}.$$

This equation can then be written as

$$E_s = E_{A_k,B_k}\left\{A_k^2 \int_{-\infty}^{+\infty} h(t-kT)^2 \cos^2(2\pi f_0 t)dt + B_k^2 \int_{-\infty}^{+\infty} h(t-kT)^2 \sin^2(2\pi f_0 t)dt - 2A_kB_k \int_{-\infty}^{+\infty} h(t-kT)^2 \cos(2\pi f_0 t)\sin(2\pi f_0 t)dt\right\},$$

which is equivalent to

$$E_s = E_{A_k,B_k}\{A_k^2\} \int_{-\infty}^{+\infty} h(t-kT)^2 \cos^2(2\pi f_0 t)dt + E_{A_k,B_k}\{B_k^2\} \int_{-\infty}^{+\infty} h(t-kT)^2 \sin^2(2\pi f_0 t)dt - 2E_{A_k,B_k}\{A_kB_k\} \int_{-\infty}^{+\infty} h(t-kT)^2 \cos(2\pi f_0 t)\sin(2\pi f_0 t)dt.$$

By remembering our high school trigonometric expressions, we can then write

$$E_s = \frac{1}{2}E_{A_k,B_k}\{A_k^2 + B_k^2\} \int_{-\infty}^{+\infty} h(t-kT)^2 dt + \frac{1}{2}E_{A_k,B_k}\{A_k^2 - B_k^2\} \int_{-\infty}^{+\infty} h(t-kT)^2 \cos(4\pi f_0 t)dt - E_{A_k,B_k}\{A_kB_k\} \int_{-\infty}^{+\infty} h(t-kT)^2 \sin(4\pi f_0 t)dt.$$

By introducing the parameter γ defined as $\gamma = E_{A_k,B_k}\{A_k^2 + B_k^2\} = E_{C_k}\{|C_k|^2\}$ and noticing that the term $\int_{-\infty}^{+\infty} h(t-kT)^2 dt$ is simply the energy, E_h, of a pulse $h(t)$, we can write

$$E_s = \frac{\gamma}{2}E_h + \frac{1}{2}E_{A_k,B_k}\{A_k^2 - B_k^2\} \int_{-\infty}^{+\infty} h(t-kT)^2 \cos(4\pi f_0 t)dt - E_{A_k,B_k}\{A_kB_k\} \int_{-\infty}^{+\infty} h(t-kT)^2 \sin(4\pi f_0 t)dt.$$

Both terms $E_{A_k,B_k}\{A_k^2 - B_k^2\} \int_{-\infty}^{+\infty} h(t-kT)^2 \cos(4\pi f_0 t)\, dt$ and $E_{A_k,B_k}\{A_kB_k\} \int_{-\infty}^{+\infty} h(t-kT)^2 \sin(4\pi f_0 t)\, dt$ are, generally, equal to zero because, for most constellations used in practice, we have $E_{A_k,B_k}\{A_k^2 - B_k^2\} = E_{A_k,B_k}\{A_kB_k\} = 0$. This statement would ideally have to be checked for each individual constellation.

In any case, when this is not true, we can still remember that the sinusoidal waves $\cos(4\pi f_0 t)$ and $\sin(4\pi f_0 t)$ vary much faster than a pulse $h(t-kT)$. Hence, $h(t-kT)$ can be considered as a constant over the period of these waves and we can then write

$$\int_{-\infty}^{+\infty} h(t-kT)^2 \cos(4\pi f_0 t)dt \sim 0 \text{ and } \int_{-\infty}^{+\infty} h(t-kT)^2 \sin(4\pi f_0 t)dt \sim 0.$$

We conclude that the average energy allocated to each symbol C_k is expressed as $E_s = \frac{\gamma}{2}E_h$. This result means that the average SNR per symbol is given by

$$\frac{E_s}{N_0} = \frac{\gamma}{2}\frac{E_h}{N_0}.$$

Finally, over an AWGN channel, the expression of a sample C'_p is given by

$$C'_p = C_p + n_p,$$

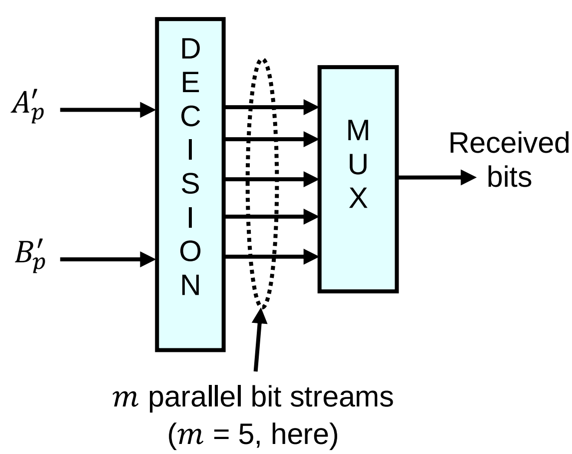
where C_p is the corresponding transmitted complex symbol, and n_p denotes a zero-mean complex Gaussian noise sample with variance $2\sigma^2 = \frac{2N_0}{E_h} = \gamma\left(\frac{E_s}{N_0}\right)^{-1}$. Note that the parameter $\gamma = E_{C_p}\left\{\left|C_p\right|^2\right\}$ can be easily computed for any constellation.

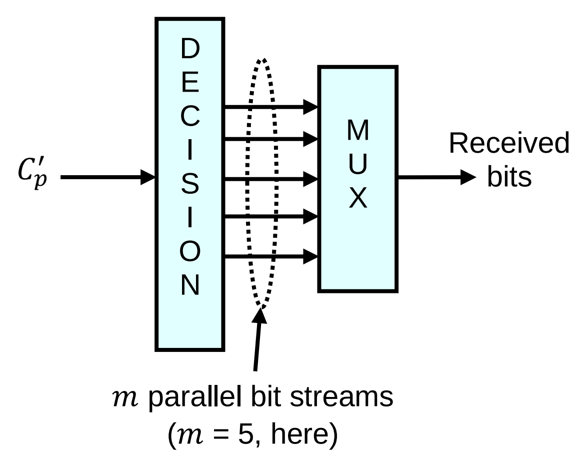
This is an important equation to be remembered. For those who do not like complex numbers, they can use the following channel model instead: the expressions of the samples A'_p and B'_p are given by

$$A'_p = A_p + n_{p,i},$$
$$B'_p = B_p + n_{p,q},$$

where A_p and B_p are the corresponding transmitted symbols, whereas $n_{p,i}$ and $n_{p,q}$ denote two independent zero-mean Gaussian noise samples with variance $\sigma^2 = \frac{\gamma}{2}\left(\frac{E_s}{N_0}\right)^{-1}$. The parameter $\gamma = E_{A_p,B_p}\{A_p^2 + B_p^2\}$ can be easily computed for any constellation.

• **Decision block**





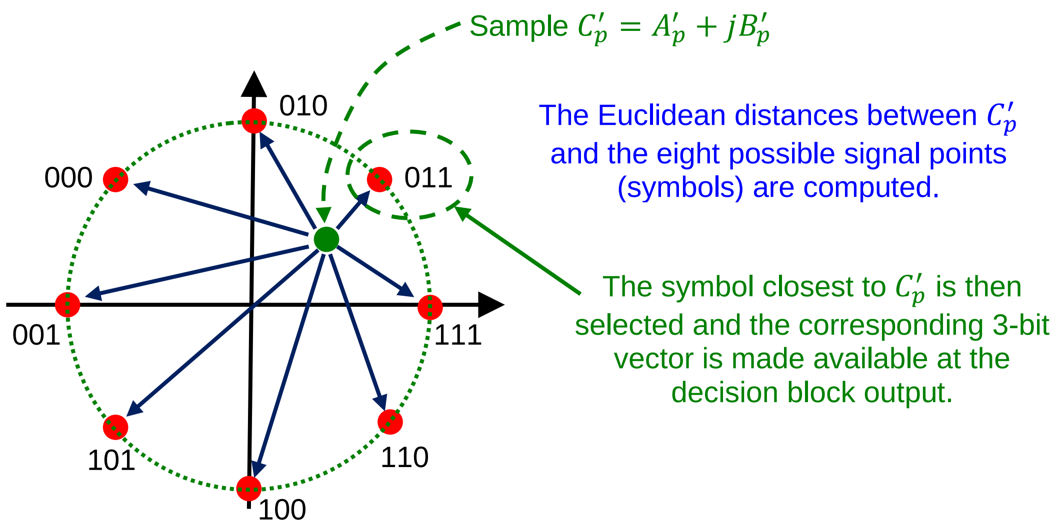
The final task performed by the receiver consists of using the sample $C'_p = A'_p + jB'_p$ in order to take a decision regarding the value of the complex symbol C_p transmitted at time pT.

Once the most likely value of C_p has been determined, it is straightforward to recover the m bits associated with it because the mapping used at the transmitter side is obviously known to the receiver.

The decision block evaluates the Euclidean distances between C'_p and the M possible symbols in the constellation, and then decides that the symbol closest to C'_p was the transmitted one.

The m-bit vector associated with this detected symbol is finally generated by the decision block.

This simple process is illustrated in the case of 8-PSK by the example shown below.



Such process, although optimal in terms of symbol detection, sometimes fails to properly detect the value of the transmitted symbol, especially when the noise samples $n_{p,i}$ and $n_{p,i}$ have large magnitudes compared to those of the transmitted symbols A_p and B_p. In other words, transmission errors occur from time to time.

In the next chapter, we are going to introduce the concept of *error probability* in a digital communication system.

9. Error Probability at the Receiver Output for AWGN Channels

In the previous chapter, we have demonstrated that, over an AWGN channel, a complex sample C'_p available at the decision block input is given by

$$C'_p = C_p + n_p,$$

where C_p is the corresponding transmitted complex symbol, and n_p denotes a zero-mean complex Gaussian noise sample with variance $2\sigma^2 = \frac{2N_0}{E_h} = \gamma\left(\frac{E_s}{N_0}\right)^{-1}$.

The error performance of a digital communications scheme can be assessed by evaluating, via calculations or computer simulations, the *symbol error probability* and/or the *bit error probability* at the receiver output.

Let us first focus our attention on the symbol error probability, P_{es}, defined as the probability that the decision block chooses the wrong symbol:

$$P_{es} = \Pr\{\widetilde{C_p} \neq C_p\},$$

where $\widetilde{C_p}$ is the symbol detected by the decision block and C_p is the transmitted symbol.

• **Expression of the symbol error probability for any modulation scheme**

Let us drop the time index p since the time is not relevant throughout the following calculations. We replace it with an index i which indicates a particular symbol in the constellation.

Thus, from now on, we are going to adopt the following notations:

- $C = A + jB$ is the transmitted complex symbol;
- $C' = A' + jB' = (A + n_i) + j(B + n_q)$ is the estimate of $C = A + jB$ available at the decision block input;
- $\tilde{C}$ is the symbol detected by the decision block. In other words, it is the decision taken by this decision block regarding the value of C;

- C_i, $i \in \{1, 2, \dots, M\}$, is a particular symbol in a constellation composed of M signal points.

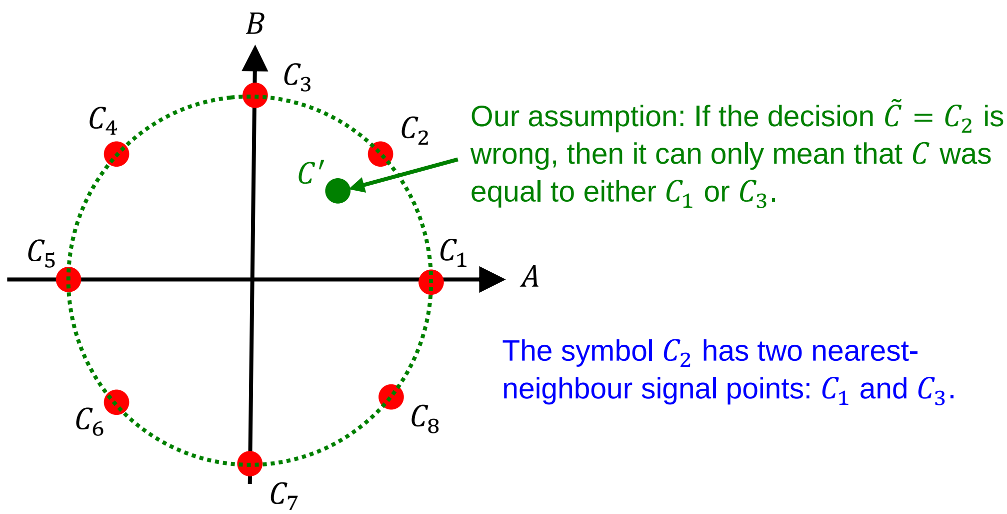
We start by noticing that

$$P_{es} = \Pr\{(C_1 \text{ transmitted} \cap C_1 \text{ not detected}) \cup (C_2 \text{ transmitted} \cap C_2 \text{ not detected}) \cup \dots \cup (C_M \text{ transmitted} \cap C_M \text{ not detected})\}.$$

As the M events $(C_i \text{ transmitted} \cap C_i \text{ not detected})$, $i \in \{1, 2, \dots, M\}$, are mutually exclusive as they can never occur at the same time, the unions can be written as a sum of probabilities:

$$P_{es} = \sum_{i=1}^{M} \Pr\{C_i \text{ transmitted} \cap C_i \text{ not detected}\}.$$

At this stage, we are going to use the following assumption which is valid at all signal-to-noise ratios (SNRs) of practical interest: when taking an erroneous decision, the decision block always chooses a symbol that is a nearest neighbour to the transmitted symbol.



With this assumption, we can write

$$P_{es} \sim \sum_{i=1}^{M} \Pr\{(C_i \text{ transmitted} \cap C_{i,1} \text{ detected}) \cup (C_i \text{ transmitted} \cap C_{i,2} \text{ detected}) \cup \dots\},$$

where $C_{i,j}$ is the j-th nearest neighbour of C_i.

Since we have once again to consider the union of events that are mutually exclusive, the previous expression can be written as a sum of probabilities:

$$P_{es} \sim \sum_{i=1}^{M} \sum_{C_j \in S_i} \Pr\{C_i \text{ transmitted} \cap C_j \text{ detected}\},$$

where S_i is the set containing all nearest neighbour symbols of C_i.

Using Bayes' rule, we obtain

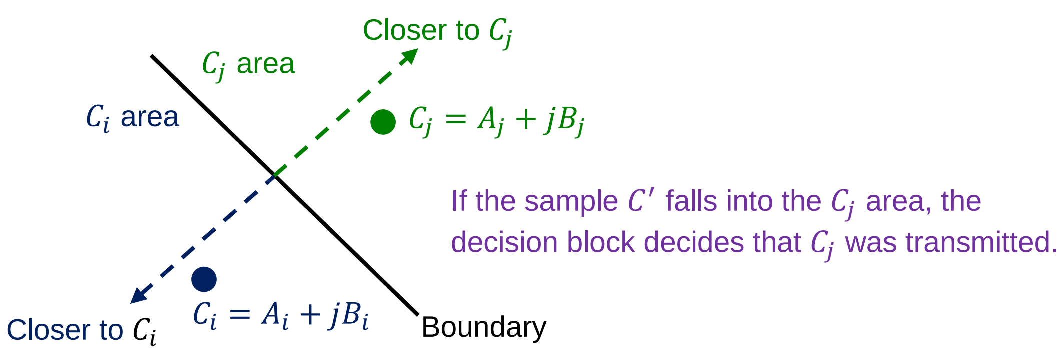
$$P_{es} \sim \sum_{i=1}^{M} \sum_{C_j \in S_i} \Pr\{C_i \to C_j\} \cdot \Pr\{C_i \text{ transmitted}\},$$

where $\Pr\{C_i \to C_j\}$ denotes the probability to detect C_j given that C_i was transmitted.

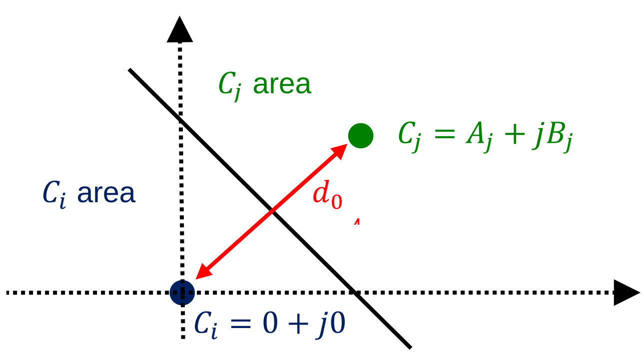
Since all symbols are transmitted with equal probabilities, we have $\Pr\{C_i \text{ transmitted}\} = \frac{1}{M}, i \in \{1, 2, \dots, M\}$, and we can thus write

$$P_{es} \sim \frac{1}{M} \cdot \sum_{i=1}^{M} \sum_{C_j \in S_i} \Pr\{C_i \to C_j\}.$$

We now need to find an expression for the term $\Pr\{C_i \to C_j\}$.



To simplify the calculations, we can assume, without any loss of generality, that $A_i = B_i = 0$.



The Euclidean distance between C_i and C_j is given by $d_0 = \sqrt{A_j^2 + B_j^2}$. Note that the distance d_0 is often referred to as the minimum distance between symbols in the constellation since C_i and C_j are two nearest neighbour symbols.

The equation of the boundary line between the C_i area and the C_j area is given by

$$y = -\left(\frac{A_j}{B_j}\right) \cdot x + \frac{d_0^2}{2B_j}.$$

If you remember well your high school mathematics, you probably know how to obtain this equation. If you do not, then you can derive it by noticing that the boundary line is composed by all the points with coordinates (x, y) that are at equal Euclidean distances from C_i and C_j. Therefore, the coordinates x and y of these points must be such that the condition

$$(x-0)^2 + (y-0)^2 = \left(x - A_j\right)^2 + \left(y - B_j\right)^2$$

is satisfied. Developing this expression leads to the expression shown above.

The probability $\Pr\{C_i \rightarrow C_j\}$ is simply the probability that the sample $C' = A' + jB'$ used by the decision block falls into the C_j area, which translates in mathematical terms as

$$\Pr\{C_i \rightarrow C_j\} = \Pr\left\{B' > -\left(\frac{A_j}{B_j}\right) \cdot A' + \frac{d_0^2}{2B_j}\right\}.$$

Since we have assumed, for simplicity's sake, that the transmitted symbol was $C_i = A_i + jB_i = 0 + j0$, we can see that $C' = A' + jB' = n_i + jn_q$, where n_i and n_q are two independent Gaussian noise samples with zero-mean and variance

$$\sigma^2 = \frac{\gamma}{2}\frac{E_s}{N_0}.$$

Thus, we have $\Pr\{C_i \rightarrow C_j\} = \Pr\left\{n_q + \left(\frac{A_j}{B_j}\right) \cdot n_i > \frac{d_0^2}{2B_j}\right\} = \Pr\left\{n_{i,q} > \frac{d_0^2}{2B_j}\right\}$.

The noise sample $n_{i,q} = n_q + \left(\frac{A_j}{B_j}\right) \cdot n_i$ is also a Gaussian sample with a mean given by

$$m = E\{n_{i,q}\} = E\left\{n_q + \left(\frac{A_j}{B_j}\right) \cdot n_i\right\} = E\{n_q\} + \left(\frac{A_j}{B_j}\right) \cdot E\{n_i\} = 0$$

and a variance equal to

$$E\{n_{i,q}^2\} = E\left\{n_q^2 + 2n_q n_i \left(\frac{A_j}{B_j}\right) + \left(\frac{A_j}{B_j}\right)^2 \cdot n_i^2\right\}.$$

Developing this expression leads to

$$E\{n_{i,q}^2\} = E\{n_q^2\} + 2 \cdot \left(\frac{A_j}{B_j}\right) \cdot E\{n_q\} \cdot E\{n_i\} + \left(\frac{A_j}{B_j}\right)^2 \cdot E\{n_i^2\},$$

which yields

$$E\{n_{i,q}^2\} = \sigma^2 + \left(\frac{A_j}{B_j}\right)^2 \sigma^2 = \sigma^2 \cdot \left[1 + \left(\frac{A_j}{B_j}\right)^2\right] = \sigma^2 \frac{d_0^2}{B_j^2}.$$

The probability density function (PDF) of the Gaussian noise sample $n_{i,q}$ is given by

$$P_{n_{i,q}}(x) = \frac{1}{\sqrt{2\pi\sigma^2 \frac{d_0^2}{B_j^2}}} \cdot e^{-\frac{x^2}{2\sigma^2 \frac{d_0^2}{B_j^2}}} = \frac{B_j}{\sqrt{2\pi\sigma^2 d_0^2}} \cdot e^{-\frac{B_j^2 x^2}{2\sigma^2 d_0^2}},$$

with $\sigma^2 = \frac{\gamma}{2}\left(\frac{E_S}{N_0}\right)^{-1}$.

By applying a basic property of PDFs which says that $\Pr\{a < n_{i,q} < b\} = \int_a^b P_{n_{i,q}}(x)dx$, it can be shown that

$$\Pr\{C_i \to C_j\} = \frac{B_j}{\sqrt{2\pi\sigma^2 d_0^2}} \cdot \int_{\frac{d_0^2}{2B_j}}^{+\infty} e^{-\frac{B_j^2 x^2}{2\sigma^2 d_0^2}} dx.$$

By introducing the *complementary error function* defined as $\operatorname{erfc}(x) = \frac{2}{\sqrt{\pi}} \cdot \int_x^{+\infty} e^{-u^2}\, du$, and performing the change of variable $u = x \cdot \sqrt{\frac{B_j^2}{2\sigma^2 d_0^2}}$, we obtain the expression

$$\Pr\{C_i \to C_j\} = \frac{B_j}{\sqrt{2\pi\sigma^2 d_0^2}} \cdot \sqrt{\frac{2\sigma^2 d_0^2}{B_j^2}} \int_{\frac{d_0^2}{2B_j}\sqrt{\frac{B_j^2}{2\sigma^2 d_0^2}}}^{+\infty} e^{-u^2} du = \frac{1}{2} \cdot \mathrm{erfc}\left(\sqrt{\frac{d_0^2}{8\sigma^2}}\right).$$

A table giving the value of $\mathrm{erfc}(x)$ as a function of x is shown in the next page.

By using $\sigma^2 = \frac{\gamma}{2}\left(\frac{E_s}{N_0}\right)^{-1}$, we finally obtain the expression of $\Pr\{C_i \to C_j\}$:

$$\Pr\{C_i \to C_j\} = \frac{1}{2} \cdot \mathrm{erfc}\left(\sqrt{\frac{d_0^2}{4\gamma}\frac{E_s}{N_0}}\right).$$

The symbol error probability can then be approximated as

$$P_{es} \sim \frac{1}{2M} \cdot \sum_{i=1}^{M} \sum_{C_j \in S_i} \mathrm{erfc}\left(\sqrt{\frac{d_0^2}{4\gamma}\frac{E_s}{N_0}}\right).$$

Remarkably, the term $\mathrm{erfc}\left(\sqrt{\frac{d_0^2}{4\gamma}\frac{E_s}{N_0}}\right)$ does not depend on the actual values of symbols C_i and C_j, but on the Euclidean distance between these symbols.

For all symbols $C_j \in S_i$ the terms $\mathrm{erfc}\left(\sqrt{\frac{d_0^2}{4\gamma}\frac{E_s}{N_0}}\right)$ are identical since all these symbols are at the same Euclidean distance d_0 from C_i. Therefore, we can write

$$P_{es} \sim \frac{1}{2M} \cdot \sum_{i=1}^{M} \mathrm{erfc}\left(\sqrt{\frac{d_0^2}{4\gamma}\frac{E_s}{N_0}}\right) \cdot \sum_{C_j \in S_i} 1 = \frac{1}{2M} \cdot \sum_{i=1}^{M} \mathrm{erfc}\left(\sqrt{\frac{d_0^2}{4\gamma}\frac{E_s}{N_0}}\right) \cdot N_i,$$

where N_i is the number of nearest neighbour symbols of C_i.

This expression can be further simplified as follows:

$$P_{es} \sim \frac{1}{2M} \cdot \mathrm{erfc}\left(\sqrt{\frac{d_0^2}{4\gamma}\frac{E_s}{N_0}}\right) \cdot \sum_{i=1}^{M} N_i,$$

since the term $\mathrm{erfc}\left(\sqrt{\frac{d_0^2}{4\gamma}\frac{E_s}{N_0}}\right)$ does not depend on the symbol C_i under consideration.

		Complementary Error Function Table											
x	**erfc(x)**	**x**	**erfc(x)**	**x**	**erfc(x)**	**x**	**erfc(x)**	**x**	**erfc(x)**	**x**	**erfc(x)**	**x**	**erfc(x)**
0	1.000000	0.5	0.479500	1	0.157299	1.5	0.033895	2	0.004678	2.5	0.000407	3	0.00002209
0.01	0.988717	0.51	0.470756	1.01	0.153190	1.51	0.032723	2.01	0.004475	2.51	0.000386	3.01	0.00002074
0.02	0.977435	0.52	0.462101	1.02	0.149162	1.52	0.031587	2.02	0.004281	2.52	0.000365	3.02	0.00001947
0.03	0.966159	0.53	0.453536	1.03	0.145216	1.53	0.030484	2.03	0.004094	2.53	0.000346	3.03	0.00001827
0.04	0.954889	0.54	0.445061	1.04	0.141350	1.54	0.029414	2.04	0.003914	2.54	0.000328	3.04	0.00001714
0.05	0.943628	0.55	0.436677	1.05	0.137564	1.55	0.028377	2.05	0.003742	2.55	0.000311	3.05	0.00001608
0.06	0.932378	0.56	0.428384	1.06	0.133856	1.56	0.027372	2.06	0.003577	2.56	0.000294	3.06	0.00001508
0.07	0.921142	0.57	0.420184	1.07	0.130227	1.57	0.026397	2.07	0.003418	2.57	0.000278	3.07	0.00001414
0.08	0.909922	0.58	0.412077	1.08	0.126674	1.58	0.025453	2.08	0.003266	2.58	0.000264	3.08	0.00001326
0.09	0.898719	0.59	0.404064	1.09	0.123197	1.59	0.024538	2.09	0.003120	2.59	0.000249	3.09	0.00001243
0.1	0.887537	0.6	0.396144	1.1	0.119795	1.6	0.023652	2.1	0.002979	2.6	0.000236	3.1	0.00001165
0.11	0.876377	0.61	0.388319	1.11	0.116467	1.61	0.022793	2.11	0.002845	2.61	0.000223	3.11	0.00001092
0.12	0.865242	0.62	0.380589	1.12	0.113212	1.62	0.021962	2.12	0.002716	2.62	0.000211	3.12	0.00001023
0.13	0.854133	0.63	0.372954	1.13	0.110029	1.63	0.021157	2.13	0.002593	2.63	0.000200	3.13	0.00000958
0.14	0.843053	0.64	0.365414	1.14	0.106918	1.64	0.020378	2.14	0.002475	2.64	0.000189	3.14	0.00000897
0.15	0.832004	0.65	0.357971	1.15	0.103876	1.65	0.019624	2.15	0.002361	2.65	0.000178	3.15	0.00000840
0.16	0.820988	0.66	0.350623	1.16	0.100904	1.66	0.018895	2.16	0.002253	2.66	0.000169	3.16	0.00000786
0.17	0.810008	0.67	0.343372	1.17	0.098000	1.67	0.018190	2.17	0.002149	2.67	0.000159	3.17	0.00000736
0.18	0.799064	0.68	0.336218	1.18	0.095163	1.68	0.017507	2.18	0.002049	2.68	0.000151	3.18	0.00000689
0.19	0.788160	0.69	0.329160	1.19	0.092392	1.69	0.016847	2.19	0.001954	2.69	0.000142	3.19	0.00000644
0.2	0.777297	0.7	0.322199	1.2	0.089686	1.7	0.016210	2.2	0.001863	2.7	0.000134	3.2	0.00000603
0.21	0.766478	0.71	0.315335	1.21	0.087045	1.71	0.015593	2.21	0.001776	2.71	0.000127	3.21	0.00000564
0.22	0.755704	0.72	0.308567	1.22	0.084466	1.72	0.014997	2.22	0.001692	2.72	0.000120	3.22	0.00000527
0.23	0.744977	0.73	0.301896	1.23	0.081950	1.73	0.014422	2.23	0.001612	2.73	0.000113	3.23	0.00000493
0.24	0.734300	0.74	0.295322	1.24	0.079495	1.74	0.013865	2.24	0.001536	2.74	0.000107	3.24	0.00000460
0.25	0.723674	0.75	0.288845	1.25	0.077100	1.75	0.013328	2.25	0.001463	2.75	0.000101	3.25	0.00000430
0.26	0.713100	0.76	0.282463	1.26	0.074764	1.76	0.012810	2.26	0.001393	2.76	0.000095	3.26	0.00000402
0.27	0.702582	0.77	0.276179	1.27	0.072486	1.77	0.012309	2.27	0.001326	2.77	0.000090	3.27	0.00000376
0.28	0.692120	0.78	0.269990	1.28	0.070266	1.78	0.011826	2.28	0.001262	2.78	0.000084	3.28	0.00000351
0.29	0.681717	0.79	0.263897	1.29	0.068101	1.79	0.011359	2.29	0.001201	2.79	0.000080	3.29	0.00000328
0.3	0.671373	0.8	0.257899	1.3	0.065992	1.8	0.010909	2.3	0.001143	2.8	0.000075	3.3	0.00000306
0.31	0.661092	0.81	0.251997	1.31	0.063937	1.81	0.010475	2.31	0.001088	2.81	0.000071	3.31	0.00000285
0.32	0.650874	0.82	0.246189	1.32	0.061935	1.82	0.010057	2.32	0.001034	2.82	0.000067	3.32	0.00000266
0.33	0.640721	0.83	0.240476	1.33	0.059985	1.83	0.009653	2.33	0.000984	2.83	0.000063	3.33	0.00000249
0.34	0.630635	0.84	0.234857	1.34	0.058086	1.84	0.009264	2.34	0.000935	2.84	0.000059	3.34	0.00000232
0.35	0.620618	0.85	0.229332	1.35	0.056238	1.85	0.008889	2.35	0.000889	2.85	0.000056	3.35	0.00000216
0.36	0.610670	0.86	0.223900	1.36	0.054439	1.86	0.008528	2.36	0.000845	2.86	0.000052	3.36	0.00000202
0.37	0.600794	0.87	0.218560	1.37	0.052688	1.87	0.008179	2.37	0.000803	2.87	0.000049	3.37	0.00000188
0.38	0.590991	0.88	0.213313	1.38	0.050984	1.88	0.007844	2.38	0.000763	2.88	0.000046	3.38	0.00000175
0.39	0.581261	0.89	0.208157	1.39	0.049327	1.89	0.007521	2.39	0.000725	2.89	0.000044	3.39	0.00000163
0.4	0.571608	0.9	0.203092	1.4	0.047715	1.9	0.007210	2.4	0.000689	2.9	0.000041	3.4	0.00000152
0.41	0.562031	0.91	0.198117	1.41	0.046148	1.91	0.006910	2.41	0.000654	2.91	0.000039	3.41	0.00000142
0.42	0.552532	0.92	0.193232	1.42	0.044624	1.92	0.006622	2.42	0.000621	2.92	0.000036	3.42	0.00000132
0.43	0.543113	0.93	0.188437	1.43	0.043143	1.93	0.006344	2.43	0.000589	2.93	0.000034	3.43	0.00000123
0.44	0.533775	0.94	0.183729	1.44	0.041703	1.94	0.006077	2.44	0.000559	2.94	0.000032	3.44	0.00000115
0.45	0.524518	0.95	0.179109	1.45	0.040305	1.95	0.005821	2.45	0.000531	2.95	0.000030	3.45	0.00000107
0.46	0.515345	0.96	0.174576	1.46	0.038946	1.96	0.005574	2.46	0.000503	2.96	0.000028	3.46	0.00000099
0.47	0.506255	0.97	0.170130	1.47	0.037627	1.97	0.005336	2.47	0.000477	2.97	0.000027	3.47	0.00000092
0.48	0.497250	0.98	0.165769	1.48	0.036346	1.98	0.005108	2.48	0.000453	2.98	0.000025	3.48	0.00000086
0.49	0.488332	0.99	0.161492	1.49	0.035102	1.99	0.004889	2.49	0.000429	2.99	0.000024	3.49	0.00000080

If we now define the average number of nearest neighbour symbols in the constellation as

$$\bar{N} = \frac{1}{M} \cdot \sum_{i=1}^{M} N_i,$$

we finally obtain

$$P_{es} \sim \frac{\bar{N}}{2} \cdot \text{erfc}\left(\sqrt{\frac{d_0^2}{4\gamma} \frac{E_s}{N_0}}\right).$$

Such expression shows that the symbol error probability depends on several parameters:

• The average number of nearest neighbour symbols in the constellation, $\bar{N}$. The lower the value of $\bar{N}$, the lower the symbol error probability. There is clearly a higher probability of error in a constellation where signal points have more neighbours, i.e., in “denser” constellations.

• The ratio $\frac{d_0^2}{\gamma}$, with $\gamma = E_C\{|C|^2\}$. The higher the value of $\frac{d_0^2}{\gamma}$, the lower the symbol error probability. Therefore, for optimal performance in terms of error probability, we should maximise the minimal Euclidean distance d_0 between signal points, for a given value of the parameter γ.

• The signal-to-noise ratio (SNR) per symbol, $\frac{E_s}{N_0}$. The higher the SNR, the lower the symbol error probability. But, increasing the SNR comes at a cost as will be seen later.

In practice, most engineers prefer using another definition of the SNR in which E_s is replaced by E_b, where E_b is the average energy allocated to the transmission of one bit. The reason is that the fundamental unit of information in communication theory is the bit. On the other hand, the symbol cannot be considered as the reference when comparing various systems since its information content does vary depending on the constellation under consideration.

Since each symbol carries m bits, we have $E_s = mE_b$, and we can thus write

$$P_{es} \sim \frac{\bar{N}}{2} \cdot \text{erfc}\left(\sqrt{\frac{md_0^2}{4\gamma} \frac{E_b}{N_0}}\right),$$

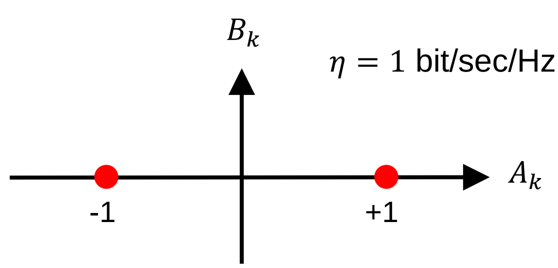
where $\frac{E_b}{N_0}$ designates the *SNR per information bit*.

We are now going to use this generic expression to determine the expression of P_{es} for the various modulation schemes introduced earlier.

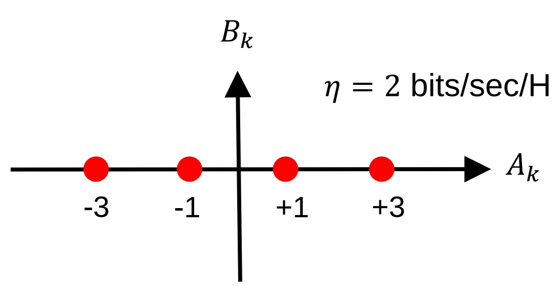
• M-state amplitude shift keying (M-ASK) modulation

For these modulation schemes, the symbol C_k is a real symbol that can take M possible values: $C_k = A_k \in \{\pm 1, \pm 3, \pm 5, \ldots\}$. By using these conventions, we have $d_0^2 = 4$, $\bar{N} = \frac{2\cdot(M-1)}{M}$, and $\gamma = \frac{M^2-1}{3}$. We can thus write $P_{es} \sim \frac{M-1}{M} \cdot \mathrm{erfc}\left(\sqrt{\frac{3m}{M^2-1}\frac{E_b}{N_0}}\right)$.

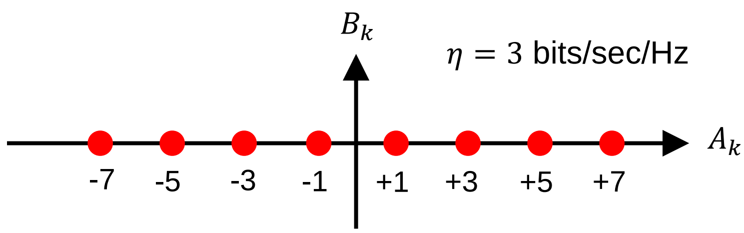
• For 2-ASK, we can write $P_{es} \sim \frac{1}{2} \cdot \mathrm{erfc}\left(\sqrt{\frac{E_b}{N_0}}\right)$. Note that, in this case, this expression does not only provide an approximation, but the exact value of the symbol error probability, i.e., we can actually write $P_{es} = \frac{1}{2} \cdot \mathrm{erfc}\left(\sqrt{\frac{E_b}{N_0}}\right)$ for 2-ASK.



• For 4-ASK, we have $P_{es} \sim \frac{3}{4} \cdot \mathrm{erfc}\left(\sqrt{\frac{2}{5}\frac{E_b}{N_0}}\right)$.



• For 8-ASK, we can write $P_{es} \sim \frac{7}{8} \cdot \operatorname{erfc}\left(\sqrt{\frac{1}{7}\frac{E_b}{N_0}}\right)$.



• **M-state phase shift keying (M-PSK) modulation**

For these modulation schemes, the complex symbol C_k has a constant modulus, i.e., C_k can be written in the form $e^{j\Phi_k}$ with $\Phi_k \in \left\{0, \frac{2\pi}{M}, \frac{4\pi}{M}, \ldots, \frac{2\pi(M-1)}{M}\right\}$. By using these conventions, we have, for $M > 2$, $d_0^2 = 4\sin^2\left(\frac{\pi}{M}\right)$, $\bar{N} = 2$, and $\gamma = 1$. We can thus write

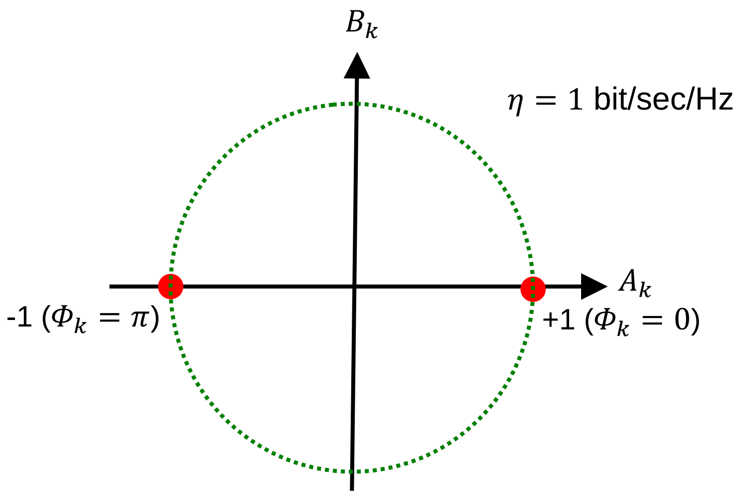
$$P_{es} \sim \operatorname{erfc}\left(\sqrt{m\sin^2\left(\frac{\pi}{M}\right)\frac{E_b}{N_0}}\right).$$

Note that, in the case $M = 2$, we have $\bar{N} = 1$ instead of $\bar{N} = 2$.

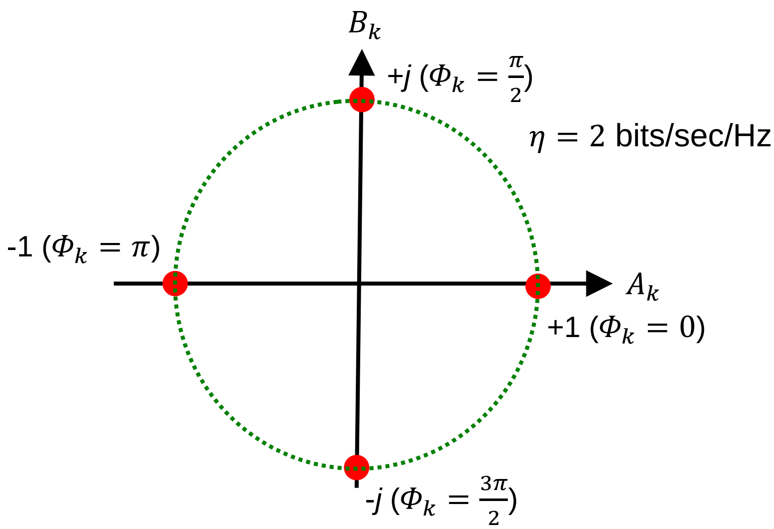
• For 2-PSK (also often referred to as BPSK), we can write $P_{es} \sim \frac{1}{2} \cdot \operatorname{erfc}\left(\sqrt{\frac{E_b}{N_0}}\right)$.

This is identical to the result obtained for 2-ASK. This was expected because 2-ASK is identical to 2-ASK. In this case, this expression does not only provide an approximation, but the exact value of the symbol error probability, i.e., we can write

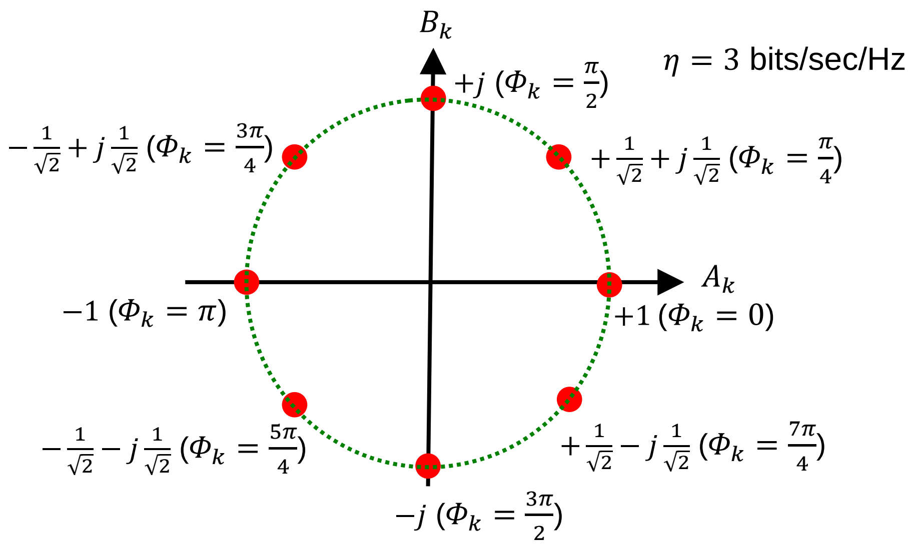
$$P_{es} = \frac{1}{2} \cdot \operatorname{erfc}\left(\sqrt{\frac{E_b}{N_0}}\right) \text{ for BPSK.}$$



• For 4-PSK (also known as QPSK), we have $P_{es} \sim \text{erfc}\left(\sqrt{\frac{E_b}{N_0}}\right)$.



• For 8-PSK, we can write $P_{es} \sim \text{erfc}\left(\sqrt{3\sin^2\left(\frac{\pi}{8}\right)\frac{E_b}{N_0}}\right)$.



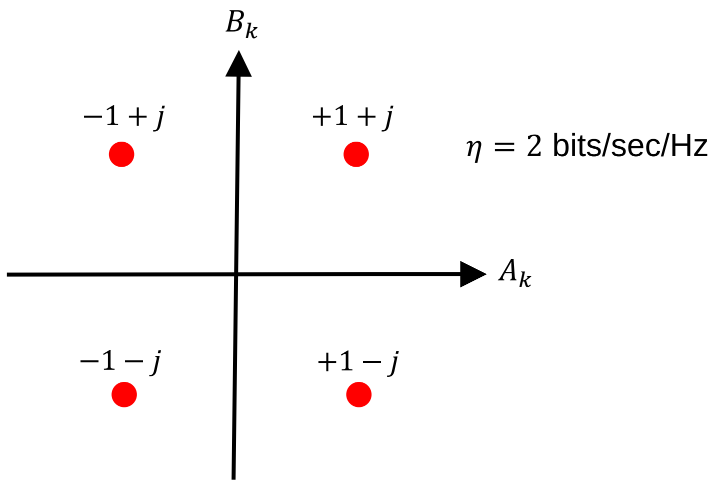
• **M-state quadrature amplitude modulation (M-QAM)**

For these modulation schemes, the symbol $C_k = A_k + jB_k$ is a complex symbol that can take M possible values, with $A_k, B_k \in \{\pm 1, \pm 3, \pm 5, \dots\}$. By using these conventions, we have $d_0^2 = 4$, $\bar{N} = 4\frac{\sqrt{M}-1}{\sqrt{M}}$, and $\gamma = \frac{2(M-1)}{3}$. We can thus write

$$P_{es} \sim 2\frac{\sqrt{M}-1}{\sqrt{M}} \cdot \text{erfc}\left(\sqrt{\frac{3m}{2(M-1)}\frac{E_b}{N_0}}\right).$$

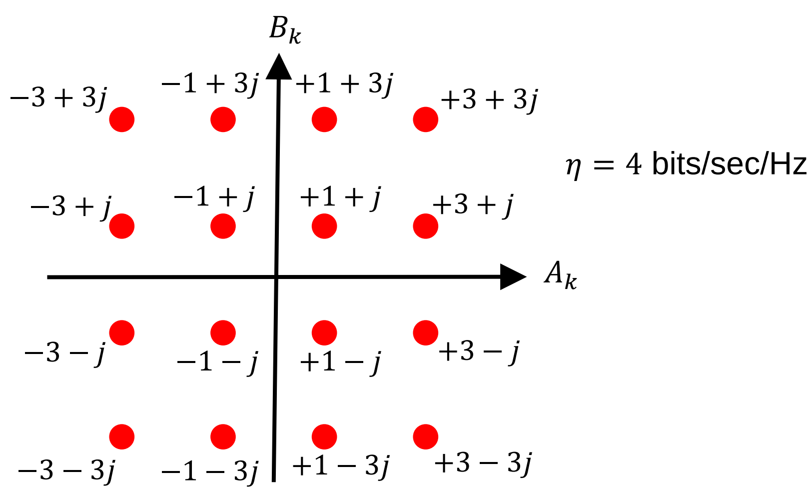
• For 4-QAM, we can write $P_{es} \sim \text{erfc}\left(\sqrt{\frac{E_b}{N_0}}\right)$.

This result is identical to that obtained for QPSK, which is not surprising as 4-QAM and QPSK are in fact identical modulation schemes.

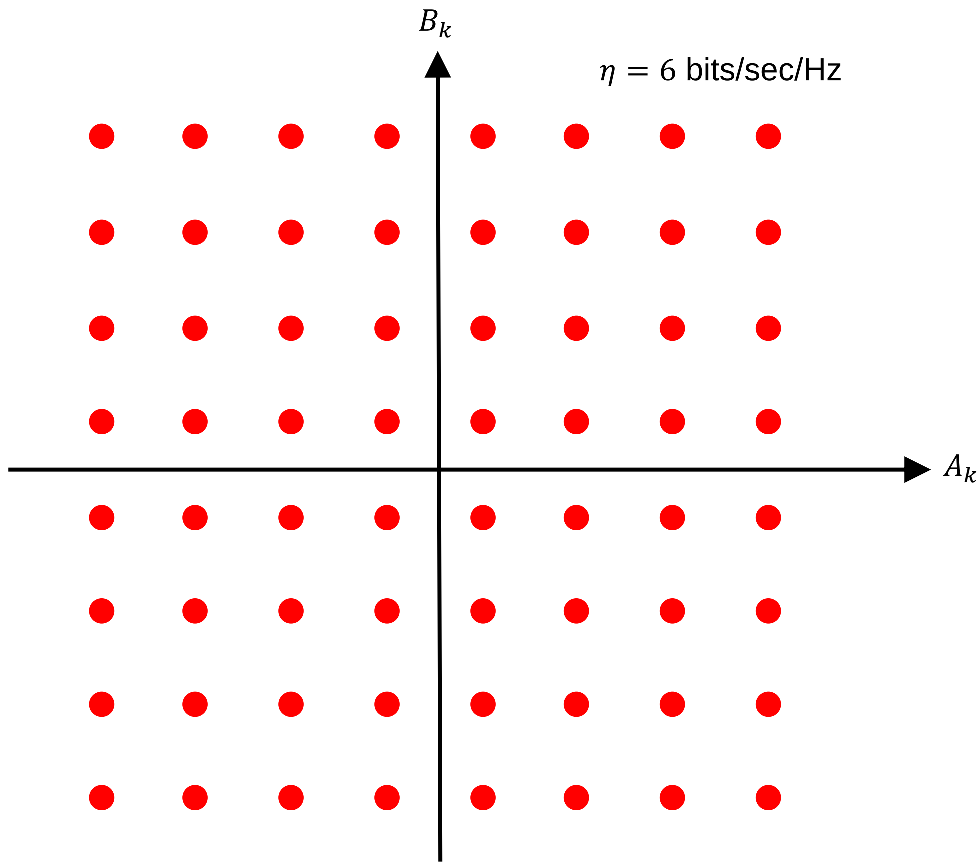


4-QAM is identical to QPSK. Do not be fooled by the $\frac{\pi}{4}$ phase difference between both constellations, which does not change anything in practice, and the different symbol values, which are only a convention anyway. What matters is that the shapes of 4-QAM and QPSK are identical.

• For 16-QAM, we have $P_{es} \sim \frac{3}{2} \cdot \text{erfc}\left(\sqrt{\frac{2}{5}\frac{E_b}{N_0}}\right)$.



• **For** 64-QAM, we can write $P_{es} \sim \frac{7}{4} \cdot \text{erfc}\left(\sqrt{\frac{1}{7}\frac{E_b}{N_0}}\right)$.



- **Expression of the bit error probability for any modulation scheme**

In practice, engineers are often more interested in the bit error probability, P_{eb}, at the receiver output rather than the symbol error probability. After all, the original goal is to transmit information bits as reliably as possible.

It is very simple to obtain an expression for P_{eb} starting from

$$P_{eb} = \Pr\{\text{a bit is detected erroneously}\}.$$

Using a "trick" of probability theory, we can also write

$$P_{eb} = \Pr\{(\text{a symbol is detected erroneously}) \cap (\text{a bit is detected erroneously})\},$$

which is, according to Bayes' rule, equivalent to

$$P_{eb} = \Pr\{\text{a bit is detected erroneously}|\text{a symbol is detected erroneously}\} \cdot P_{es},$$

where $P_{es} = \Pr\{\text{a symbol is detected erroneously}\}$ denotes the symbol error probability.

If the association between bits and symbols is done using Gray mapping, we can write

$$\Pr\{\text{a bit is detected erroneously}|\text{a symbol is detected erroneously}\} = \frac{1}{m},$$

which leads to

$$P_{eb} \sim \frac{P_{es}}{m}.$$

Finally, we obtain the final expression of P_{eb} versus $\frac{E_b}{N_0}$ for all modulation schemes considered throughout these notes:

• BPSK and QPSK: $P_{eb} =/\sim \frac{1}{2} \cdot \mathrm{erfc}\left(\sqrt{\frac{E_b}{N_0}}\right)$.

These results indicate that one should use QPSK instead of BPSK as the spectral efficiency is then doubled at no cost in terms of error performance.

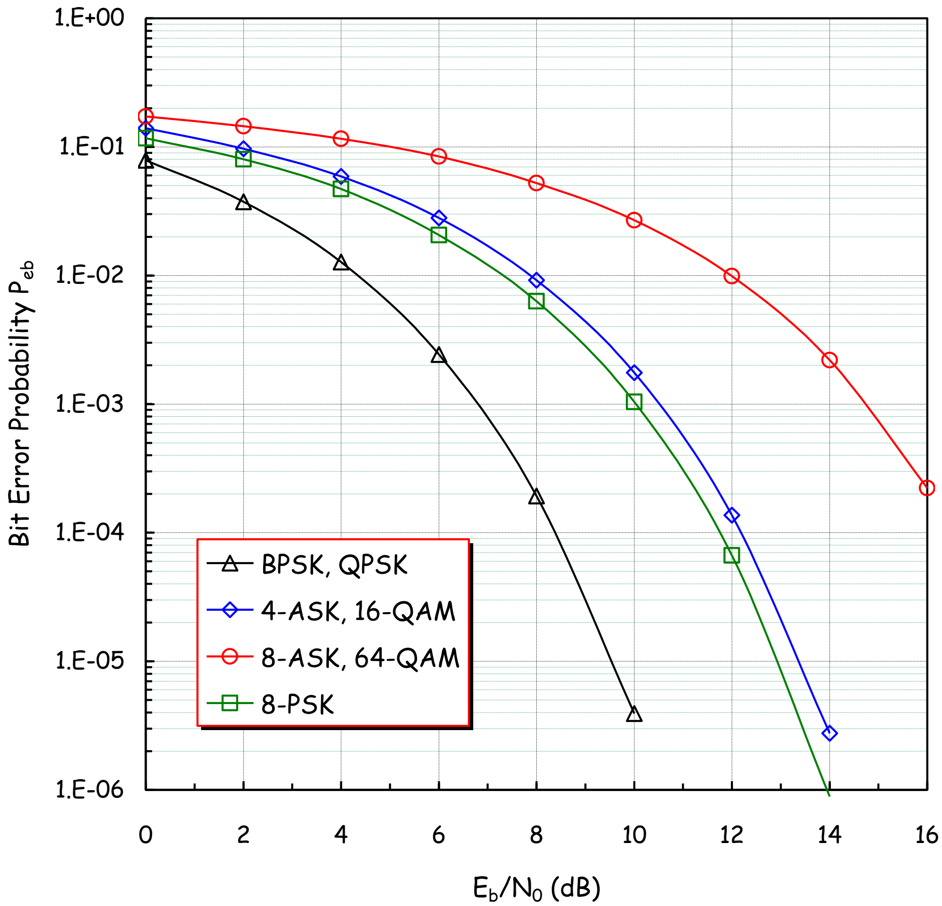
• 4-ASK and 16-QAM: $P_{eb} \sim \frac{3}{8} \cdot \mathrm{erfc}\left(\sqrt{\frac{2}{5}\frac{E_b}{N_0}}\right)$.

Once again, my advice is to use 16-QAM instead of 4-ASK for the same reason as above.

• 8-ASK and 64-QAM: $P_{eb} \sim \frac{7}{24} \cdot \mathrm{erfc}\left(\sqrt{\frac{1}{7}\frac{E_b}{N_0}}\right)$.

One should therefore use 64-QAM instead of 8-ASK for the same reason as above.

• 8-PSK: $P_{eb} \sim \frac{1}{3} \cdot \mathrm{erfc}\left(\sqrt{3\sin^2\left(\frac{\pi}{8}\right)\frac{E_b}{N_0}}\right)$.



• **Power efficiency of digital modulation schemes**

It is important to analyse the bit error probability plots properly. One must understand that our goal as communication engineers is to design systems that can operate at the *lowest possible SNR*. In fact, minimizing the operating SNR brings with it many benefits such as a longer battery life (due to the reduced energy E_b required to send each bit), lesser health hazards (due to the reduced power radiated by the transmit antenna), cheaper electronic components (due to the possibility of designing, for instance, low-noise RF amplifiers with higher noise factors), etc.

Communication schemes are generally required to achieve a certain bit error probability at the receiver output, e.g., $P_{eb} = 10^{-5}$. The challenge is to reach this target P_{eb} using the smallest possible signal-to-noise ratio $\frac{E_b}{N_0}$. A system that achieves the target P_{eb} at a low SNR is said to be *power-efficient*.

So, how do we compare the error performances, i.e., power efficiencies, of two different modulation schemes?

At low bit error probabilities, i.e., at sufficiently high SNRs, the parameter governing the error performance of a modulation scheme is (almost) exclusively the argument of the $\mathrm{erfc}(\cdot)$ function. The reason is that, as $\frac{E_b}{N_0} \to +\infty$, the slopes of all P_{eb} *versus* $\frac{E_b}{N_0}$ curves become progressively *vertical*. This essentially means that the value of the coefficient in front of the erfc(.) function becomes irrelevant as $\frac{E_b}{N_0} \to +\infty$.

As an illustration, let us compare the error performance of QPSK with that of 16-QAM. To obtain the same P_{eb} with both modulation schemes, we must have

$$\frac{1}{2} \cdot \mathrm{erfc}\left(\sqrt{\left(\frac{E_b}{N_0}\right)_{QPSK}}\right) = \frac{3}{8} \cdot \mathrm{erfc}\left(\sqrt{\frac{2}{5}\left(\frac{E_b}{N_0}\right)_{16-QAM}}\right),$$

where $\left(\frac{E_b}{N_0}\right)_{QPSK}$ and $\left(\frac{E_b}{N_0}\right)_{16-QAM}$ designate the SNRs required to obtain this P_{eb} value using QPSK and 16-QAM, respectively.

Since the desired P_{eb} value is assumed to be sufficiently low ($< 10^{-4} - 10^{-5}$), the coefficients in front of both erfc(.) functions can be ignored, and the equation thus becomes

$$\mathrm{erfc}\left(\sqrt{\left(\frac{E_b}{N_0}\right)_{QPSK}}\right) \sim \mathrm{erfc}\left(\sqrt{\frac{2}{5}\left(\frac{E_b}{N_0}\right)_{16-QAM}}\right),$$

which leads to $\left(\frac{E_b}{N_0}\right)_{QPSK} \sim \frac{2}{5}\left(\frac{E_b}{N_0}\right)_{16-QAM}$.

By using decibel notations, we obtain $\left(\frac{E_b}{N_0}\right)_{16-QAM}^{dB} - \left(\frac{E_b}{N_0}\right)_{QPSK}^{dB} = 10 \cdot \log_{10}\left(\frac{5}{2}\right) \sim 4$ dB.

This result indicates that 16-QAM must operate with a SNR that is $10 \cdot \log_{10}\left(\frac{5}{2}\right) \sim 4$ dB higher than that of QPSK in order to achieve the same bit error probability. In this case, we can simply say that QPSK is more power-efficient than 16-QAM by 4 dB.

There is however no free lunch: using QPSK rather than 16-QAM allows us to reduce the operating SNR by 4 dB, but the spectral efficiency, sometimes also referred to as *bandwidth efficiency*, is then halved ($\eta = 2$ bits/s/Hz for QPSK instead of $\eta = 4$ bits/s/Hz for 16-QAM).

There is clearly a trade-off between spectral efficiency and power efficiency. Of course, this does not come as a surprise for those among you who have heard about Claude E. Shannon and his works. If you want to know more about this topic, feel free to read my lecture notes on information theory that are available upon request.

In the same way, we could show that QPSK is $10 \cdot \log_{10}(7) \sim 8.45$ dB more power-efficient, but three times less spectral-efficient, than 64-QAM ($\eta = 2$ bits/s/Hz for QPSK and $\eta = 6$ bits/s/Hz for 64-QAM).

As for 8-PSK, it is approximately 3.6 dB less power-efficient, but one and a half times more spectral-efficient, than QPSK ($\eta = 3$ bits/s/Hz for 8-PSK).

Using QAM modulations, we can demonstrate that the SNR increase necessary for incrementing the spectral efficiency by 2 bits/sec/Hz without degrading the error probability is given by

$$10 \cdot \log_{10}\left[\frac{4M-1}{M-1} \cdot \frac{m}{m+2}\right] \text{ dB.}$$

We can thus evaluate the SNR gaps between various QAMs at high SNRs:

- Between QPSK and 16-QAM: ~ 3.98 dB;
- Between 16-QAM and 64-QAM: ~ 4.47 dB;
- Between 64-QAM and 256-QAM: ~ 4.82 dB;
- Between 256-QAM and 1024-QAM: ~ 5.06 dB.

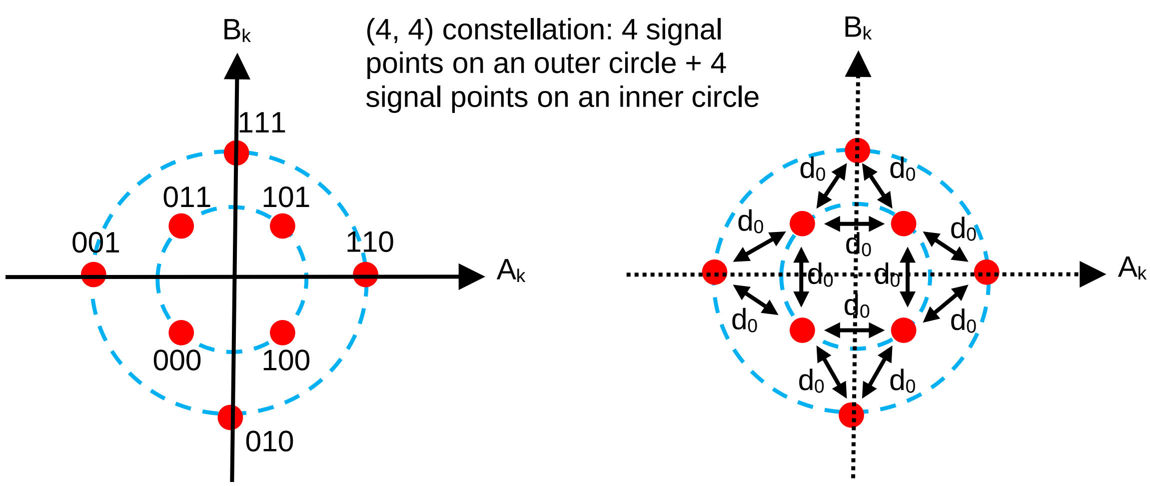
For a given spectral efficiency, we can show that an M-QAM is $-10 \cdot \log_{10}\left[\frac{2(M-1)}{3} sin^2\left(\frac{\pi}{M}\right)\right]$ dB more power-efficient than its M-PSK counterpart, thus providing us with the following SNR gap values:

- Between 4-QAM and 4-PSK: 0 dB;
- Between 16-QAM and 16-PSK: ~ 4.20 dB;
- Between 64-QAM and 64-PSK: ~ 9.95 dB.

Therefore, my advice is to always use M-QAM rather than M-PSK. Note that, in some cases, we have no choice as we can only use M-PSK. This happens, for instance, when the magnitude of the transmitted symbols must be constant in order to allow for proper operation of a non-linear power amplifier at the transmitter output.

Example ---

Consider the 3-bit/s/Hz constellation, hereafter referred to as *(4, 4) constellation*, and mapping depicted below.



The signal points are positioned symmetrically around the constellation centre in a way that the Euclidean distance between any two neighbours is equal to a constant d_0 representing the minimal distance between signal points in the constellation.

Is the mapping shown above a Gray mapping?

Show that, for the (4, 4) constellation, we have $\overline{N} = 3$ and $\gamma = d_0^2 \cdot \frac{3+\sqrt{3}}{4}$, thus leading to $P_{es} \approx \frac{3}{2} \cdot \text{erfc}\left(\sqrt{\frac{3}{3+\sqrt{3}} \cdot \frac{E_b}{N_0}}\right)$.

Also show that (4, 4) outperforms 8-PSK by 1.6 dB at high SNR.

The answers to these questions are given below.

No, this is NOT a Gray mapping because we can find several pairs of nearest neighbour signal points whose mappings differ by more than one bit. Gray mapping is actually impossible for this constellation.

To find the expression of the symbol error probability, we need to determine the value of the parameters m (number of bits per constellation signal point) and $\overline{N}$ (average number of nearest neighbour signal points). We must also determine the expression of the average energy per signal point γ as a function of d_0^2.

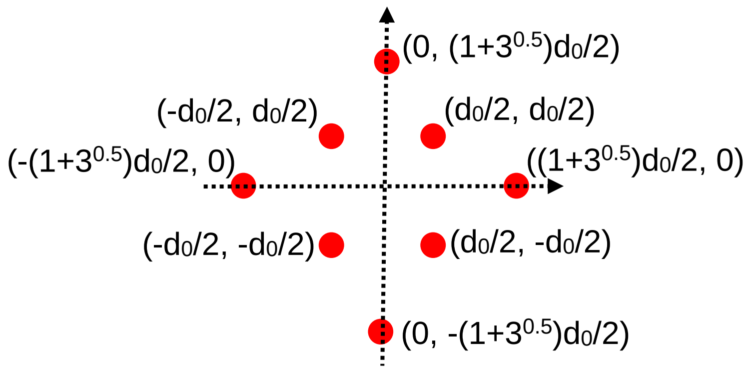
We have m = 3 since there are M = 8 signal points in the (4, 4) constellation.

A close inspection of the constellation indicates that there are four signal points with four neighbours as well as four signal points with two neighbours. Thus, we can write

$$\overline{N} = \frac{(4 \times 4) + (4 \times 2)}{8} = \frac{16+8}{8} = 3.$$

To determine the average energy per signal point γ, we need to determine the coordinates of each signal point in the (4, 4) constellation as a function of d_0^2.

The results are shown below.



We see that there are four signal points with an energy equal to $\left(\frac{d_0}{2}\right)^2 + \left(\frac{d_0}{2}\right)^2 = \frac{d_0^2}{2}$ and four signal points with an energy equal to $\frac{(1+\sqrt{3})^2 \cdot d_0^2}{4}$. Therefore, we can write

$$\gamma = \frac{4}{8}\cdot\frac{d_0^2}{2} + \frac{4}{8}\cdot\frac{(1+\sqrt{3})^2 \cdot d_0^2}{4} = \frac{2+(1+\sqrt{3})^2}{8}\cdot d_0^2 = \frac{6+2\sqrt{3}}{8}\cdot d_0^2 = \frac{3+\sqrt{3}}{4}\cdot d_0^2 .$$

The symbol error probability is given by $P_{es} \approx \frac{\overline{N}}{2}\cdot \text{erfc}\left(\sqrt{\frac{m\, d_0^2}{4\gamma}\cdot\frac{E_b}{N_0}}\right)$. Using the results shown above, we can finally write

$$P_{es} \approx \frac{3}{2}\cdot \text{erfc}\left(\sqrt{\frac{3\times 4}{4\times(3+\sqrt{3})}\cdot\frac{E_b}{N_0}}\right) = \frac{3}{2}\cdot \text{erfc}\left(\sqrt{\frac{3}{3+\sqrt{3}}\cdot\frac{E_b}{N_0}}\right).$$

We recall that the symbol error probability for 8-PSK is given by

$$P_{es} \approx \text{erfc}\left(\sqrt{3\sin^2\left(\frac{\pi}{8}\right)\cdot\frac{E_b}{N_0}}\right).$$

To compare the symbol error performance of 8-PSK with that of (4, 4) at high SNR, we only need to focus on the argument of the erfc(·) function, i.e. use the fact that both constellations achieve the same symbol error probability if

$$\mathrm{erfc}\left(\sqrt{\frac{3}{3+\sqrt{3}}\cdot\left(\frac{E_b}{N_0}\right)_{(4,4)}}\right)\approx \mathrm{erfc}\left(\sqrt{3\sin^2\left(\frac{\pi}{8}\right)\cdot\left(\frac{E_b}{N_0}\right)_{8\text{-PSK}}}\right),$$

which yields

$$\frac{3}{3+\sqrt{3}}\cdot\left(\frac{E_b}{N_0}\right)_{(4,4)}\approx 3\sin^2\left(\frac{\pi}{8}\right)\cdot\left(\frac{E_b}{N_0}\right)_{8\text{-PSK}}.$$

By using a decibel notation for the SNRs, we obtain

$$\left(\frac{E_b}{N_0}\right)_{8\text{-PSK}}-\left(\frac{E_b}{N_0}\right)_{(4,4)}\approx 10\cdot\log\left(\frac{3}{3+\sqrt{3}}\right)-10\cdot\log\left(3\sin^2\left(\frac{\pi}{8}\right)\right)$$

$$\left(\frac{E_b}{N_0}\right)_{8-\text{PSK}}-\left(\frac{E_b}{N_0}\right)_{(4,4)}\approx -10\cdot\log\left(\left(3+\sqrt{3}\right)\cdot\sin^2\left(\frac{\pi}{8}\right)\right)\approx 1.6\text{ dB.}$$

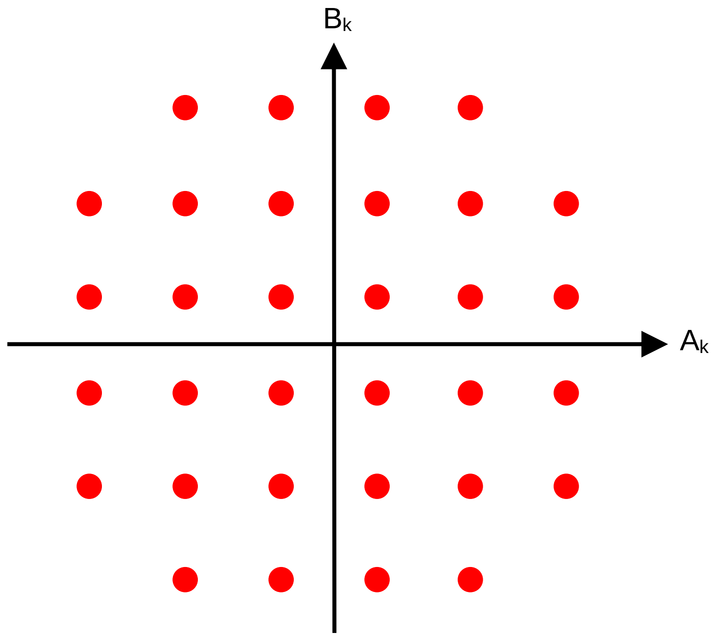
We conclude that (4, 4) is ~ 1.6 dB more power-efficient than 8-PSK at high SNRs.

Example --

Consider the 5-bit/s/Hz constellation depicted below. For such constellation, hereafter referred to as *32-QAM*, the complex symbol $C_k = A_k + jB_k$ can take 32 possible values, i.e. A_k and $B_k \in \{\pm1, \pm3, \pm5\}$ without allowing for the four particular symbols $C_k = \pm5 \pm j5$ to be generated.

Show that we have $\overline{N}=\frac{13}{4}$ and $\gamma = 5d_0^2$, thus leading to $P_{es}\approx\frac{13}{8}\cdot\mathrm{erfc}\left(\sqrt{\frac{1}{4}\cdot\frac{E_b}{N_0}}\right)$.

Also, compare the error performance of 32-QAM with that of 16-QAM at high SNRs, i.e. low symbol error probabilities.



The answers to these questions are given below.

By inspecting 32-QAM, we can see that there are

- sixteen signal points with four neighbours,
- eight signal points with three neighbours,
- and eight signal points with two neighbours.

On average, the number of nearest neighbour signal points, $\overline{N}$, is thus equal to

$$\overline{N} = \frac{(16 \times 4) + (8 \times 3) + (8 \times 2)}{32} = \frac{64 + 24 + 16}{32} = \frac{13}{4}.$$

The parameter γ is defined as $\gamma = \underset{A_k, B_k}{E}\left\{(A_k)^2 + (B_k)^2\right\}$.

By inspecting the 32-QAM constellation, we can see that there are

- four signal points for which $\gamma = 2$,
- eight signal points for which $\gamma = 10$,
- four signal points for which $\gamma = 18$,

- eigth signal points for which $\gamma = 26$,
- and eight signal points for which $\gamma = 34$.

As a result, we can determine the value of γ by taking the average of these numbers:

$$\gamma = \frac{(4\times 2)+(8\times 10)+(4\times 18)+(8\times 26)+(8\times 34)}{32} = \frac{8+80+72+208+272}{32} = 20.$$

The symbol error probability is given by $P_{es} \approx \frac{\overline{N}}{2}\cdot \mathrm{erfc}\left(\sqrt{\frac{m\, d_0^2}{4\gamma}\cdot\frac{E_b}{N_0}}\right)$.

Using the results shown above, we can finally write

$$P_{es} \approx \frac{13}{4\times 2}\cdot \mathrm{erfc}\left(\sqrt{\frac{5\times 4}{4\times 20}\cdot\frac{E_b}{N_0}}\right) = \frac{13}{8}\cdot \mathrm{erfc}\left(\sqrt{\frac{1}{4}\cdot\frac{E_b}{N_0}}\right).$$

The symbol error probability for 16-QAM is given by $P_{es} \approx \frac{3}{2}\cdot \mathrm{erfc}\left(\sqrt{\frac{2}{5}\cdot\frac{E_b}{N_0}}\right)$.

To compare the symbol error performance of 32-QAM and 16-QAM at high SNR, we only need to focus on the argument in the erfc(·) function, i.e. use the fact that

$$P_{es}(16\text{-QAM}) \approx P_{es}(32\text{-QAM}) \text{ if } \mathrm{erfc}\left(\sqrt{\frac{2}{5}\cdot\left(\frac{E_b}{N_0}\right)_{16\text{-QAM}}}\right) \approx \mathrm{erfc}\left(\sqrt{\frac{1}{4}\cdot\left(\frac{E_b}{N_0}\right)_{32\text{-QAM}}}\right),$$

which yields

$$\frac{2}{5}\cdot\left(\frac{E_b}{N_0}\right)_{16\text{-QAM}} \approx \frac{1}{4}\cdot\left(\frac{E_b}{N_0}\right)_{32\text{-QAM}}.$$

By using a decibel notation for the SNRs, we finally obtain

$$\left(\frac{E_b}{N_0}\right)_{32\text{-QAM}} - \left(\frac{E_b}{N_0}\right)_{16\text{-QAM}} \approx 10\cdot\log_{10}\left(\frac{8}{5}\right) \approx 2.04 \text{ dB}.$$

This result means that 16-QAM is ~ 2.04 dB more power-efficient than 32-QAM.
